

博 士 論 文

英語文章読解項目における錯乱枝の選択率

—— 受検者の典型的な誤答に着目して ——

寺 尾 尚 大



# 目次

論文要約	1
<b>第1章 序 論</b>	<b>5</b>
1.1 能力の測定におけるテスト項目の品質の担保	5
1.1.1 テスト項目の品質	8
1.1.2 わが国におけるテスト項目の品質管理の現状	9
1.1.3 妥当性の証拠を得ることのできるテスト開発	10
1.1.4 項目作成における認知心理学の役割と本論文の立ち位置	13
1.2 多枝選択式項目の特徴と錯乱枝の機能	15
1.2.1 多枝選択式項目の特徴	16
1.2.2 多枝選択式項目の長所と短所	17
1.2.3 多枝選択式項目における選択枝の役割	18
1.2.4 錯乱枝の機能	20
1.3 英語文章読解能力とその認知モデル	22
1.3.1 英語文章読解能力を取り上げる意義	22
1.3.2 英語文章読解能力の2つの側面	23
1.3.3 英語文章中の局所的な情報処理に関する能力	25
1.3.4 英語文章の要点を把握する能力	26
1.3.5 英語文章読解項目における錯乱枝についての先行研究とその課題	27
1.4 本論文の目的と構成	29
1.4.1 本論文の目的	29
1.4.2 本論文の構成と各章の概要	29
<b>第2章 英語文章読解項目における錯乱枝の基礎的研究</b>	<b>33</b>
2.1 研究の背景	33
2.2 研究1：多枝選択式問題における錯乱枝の特徴の把握	34
2.2.1 方 法	34
2.2.2 結果と考察	37

2.3	研究2：受検者の能力群ごとにみた錯乱枝の選択率	41
2.3.1	方 法	41
2.3.2	結 果	45
2.3.3	考 察	61
2.4	第2章のまとめ	66
<b>第3章</b>	<b>英語文章読解テスト解答時の情報処理プロセスに基づく錯乱枝開発</b>	<b>69</b>
3.1	研究の背景	69
3.2	研究3：キーセンテンスの特徴に関する検討	72
3.2.1	方 法	72
3.2.2	結果と考察	76
3.3	研究4：キーセンテンスの特徴と錯乱枝の選択率との関連	76
3.3.1	方 法	76
3.3.2	結 果	80
3.3.3	考 察	88
3.4	研究5：錯乱枝における使用語と設問タイプが錯乱枝の選択率に及ぼす影響	90
3.4.1	方 法	90
3.4.2	結 果	94
3.4.3	考 察	101
3.5	第3章のまとめ	103
<b>第4章</b>	<b>英語文章の要点を把握する能力を測る問題の錯乱枝開発</b>	<b>107</b>
4.1	研究の背景	107
4.2	研究6：英語文章要約問題における典型的な誤答パターンの検討	109
4.2.1	方 法	109
4.2.2	結果と考察	118
4.3	研究7：英語文章の要点を把握する問題における錯乱枝の検討	123
4.3.1	方 法	123
4.3.2	結 果	128
4.3.3	考 察	133
4.4	第4章のまとめ	136
<b>第5章</b>	<b>総括的討論</b>	<b>139</b>
5.1	本論文の知見のまとめ	139
5.2	本論文における研究の意義と今後の課題	142
5.2.1	本論文における研究の意義	142

5.2.2	今後の課題 . . . . .	144
5.3	今後の展望 . . . . .	146
5.3.1	認知診断を目的としたテストにおける項目作成 . . . . .	146
5.3.2	自動項目生成の基礎的知見としての利用 . . . . .	147
5.4	結語 . . . . .	148
	引用文献	148
	謝辞	155



# 表目次

1.1	TOEFL iBT における妥当性の命題と対応する証拠 . . . . .	12
2.1	研究 1 で使用した調査冊子・文章・設問数一覧 . . . . .	35
2.2	研究 2 で扱う錯乱枝の要因・水準とその定義 . . . . .	40
2.3	研究 2 で使用した文章・設問一覧 . . . . .	42
2.4	テスト冊子の構成 . . . . .	44
2.5	項目分析の結果 . . . . .	46
2.6	冊子ごとの得点の記述統計量 . . . . .	48
2.7	分析 1 で扱う分割表 . . . . .	51
2.8	分析 1 におけるオッズ比の推定結果 . . . . .	54
2.9	分析 2-1 で扱う分割表 . . . . .	55
2.10	分析 2-2 で扱う分割表 . . . . .	55
2.11	分析 2-1 におけるオッズ比の推定結果 . . . . .	56
2.12	分析 2-2 におけるオッズ比の推定結果 . . . . .	57
2.13	分析 3-1 のオッズ比の推定結果 . . . . .	59
2.14	分析 3-2 におけるオッズ比の推定結果 . . . . .	60
3.1	研究 3 で用いた文章のリーダビリティ . . . . .	73
3.2	研究 4 で用いた文章一覧 . . . . .	77
3.3	項目分析の結果 . . . . .	80
3.4	冊子ごとの得点の記述統計量 . . . . .	83
3.5	分析 4-1 におけるオッズ比の推定結果 . . . . .	87
3.6	分析 4-2 におけるオッズ比の推定結果 . . . . .	87
3.7	研究 5 で用いた文章一覧 . . . . .	90
3.8	項目分析の結果 . . . . .	95
3.9	冊子ごとの得点の記述統計量 . . . . .	97
3.10	分析 5-1 におけるオッズ比の推定結果 . . . . .	100
3.11	分析 5-2 におけるオッズ比の推定結果 . . . . .	101

4.1	研究6で用いた文章一覧	110
4.2	最終版採点基準の例(冊子1・第2問)	113
4.3	解答カテゴリの一覧	115
4.4	指定段落要約問題・削除における各カテゴリの度数表	119
4.5	指定段落要約問題・一般化における各カテゴリの度数表	120
4.6	指定段落要約問題・統合における各カテゴリの度数表	120
4.7	文章全体要約問題における各カテゴリの度数表	122
4.8	研究7で用いた文章一覧	124
4.9	項目の仕様を記した一覧表の例(冊子1・第2問)	125
4.10	項目分析の結果(研究7)	128
4.11	冊子ごとの得点の記述統計量	129
4.12	分析7-1におけるオッズ比の推定結果	132
4.13	分析7-2におけるオッズ比の推定結果	133
4.14	分析7-3におけるオッズ比の推定結果	133



# 目次

1.1	Wilson (2005) における項目応答の 4 要素モデル . . . . .	6
1.2	構築一統合モデルにおける 4 種類の読解(犬塚 (2006) をもとに著者が作成)と本論文が測定対象とする能力 . . . . .	24
1.3	Embretson & Wetzel (1987) のモデル . . . . .	25
1.4	本論文の構成 . . . . .	30
2.1	研究 1 で使用した調査冊子の例 (設問・選択枝・コメント欄) . . . . .	37
2.2	誤答の理由に関するコメントの集計結果 . . . . .	38
2.3	研究 2 で使用した設問の例 . . . . .	43
2.4	受検者の得点分布 . . . . .	48
2.5	受検者の潜在特性値の推定値のヒストグラム . . . . .	49
2.6	文章中の記述の有無 (設問種別 A) のトレースライン . . . . .	53
2.7	対概念の使用 (設問種別 B) のトレースライン . . . . .	55
2.8	言及対象の取り違え (設問種別 C) のトレースライン . . . . .	58
3.1	研究 3 で使用した調査冊子の例 . . . . .	75
3.2	研究 4 における設問と対応するキーセンテンスの例 . . . . .	78
3.3	受検者の得点分布 (研究 4) . . . . .	83
3.4	受検者の潜在特性値の EAP 推定値のヒストグラム (研究 4) . . . . .	84
3.5	条件ごとにみたトレースライン (研究 4) . . . . .	86
3.6	下位レベル設問 (研究 5) . . . . .	93
3.7	上位レベル設問 (研究 5) . . . . .	94
3.8	受検者の得点分布 (研究 5) . . . . .	97
3.9	受検者の潜在特性値の EAP 推定値のヒストグラム . . . . .	98
3.10	設問タイプ・条件ごとにみたトレースライン (研究 5) . . . . .	99
4.1	受検者の得点分布 (研究 7) . . . . .	129
4.2	受検者の潜在特性値の EAP 推定値のヒストグラム (研究 7) . . . . .	130

4.3	研究7におけるトレースライン（第2問） . . . . .	132
-----	-------------------------------	-----

## 論文要約

本論文の目的は、多枝選択式の英語文章読解項目において、受検者の典型的な誤答を反映した選択枝（錯乱枝）の選択率を能力群別に検討することである。海外のテストでは、作成されたテスト項目に対して予備テストが実施され、困難度・識別力・信頼性など統計的な基準を満たしているもののみを本番のテストで使用するという方法でテストの品質管理を行っている。一方わが国では、テスト項目を公開する文化などの影響から予備テストを実施することが困難な場合があり、テスト項目が統計的な基準を満たしているかどうかの確認を行うことができない状況が多く存在する。この点は、わが国におけるテストの品質管理の大きな問題点であると言える。

予備テストを実施できない状況下では、可能な限り統計的な基準を満たせるよう、テスト項目作成の段階での工夫が必要となる。大規模に実施され、かつ個人の処遇を左右するようなハイスタークスのテストにおいてよく用いられる多枝選択式項目の場合、選択枝、特に誤答選択枝の作成には多大な時間と労力がかかると言われるが、受検者の典型的な誤答を反映させた誤答選択枝（錯乱枝）を作成することができれば、妥当性の向上とともに、項目の統計的基準を満たすことにつながる可能性がある。多枝選択式項目の品質を担保する上では、いかに錯乱枝に受検者の典型的な誤答を反映させるかということが重要な課題となる。そこで本論文は、妥当性の高い測定を可能にすることを目指し、英語文章読解能力を例に、認知心理学・第二言語習得の領域の知見を踏まえて受検者の解答データから典型的な誤答の特徴を抽出した上で、それらを反映させた錯乱枝の選択率の様相について、能力群による差異という観点から検討を行った。

第1章では、1.1節において、能力を測定する際のテスト項目の品質の担保について、わが国の現状を踏まえながら述べ、1.2節において、多枝選択式項目の特徴・長所と短所、錯乱枝の機能等に関する知見を整理した。1.3節では、英語文章読解能力について、認知心理学における理論をレビューした。1.4節では、本論文の目的と構成、各章の概要について述べた。

第2章では、過去に実施された英語文章読解テストの誤答選択枝に含まれる誤答の特徴を抽出し、その特徴を含んだ錯乱枝の選択率を能力群別に検討する実験を行った。2.1節では研究の背景として、英語文章読解項目の錯乱枝に関する研究知見が乏しい点を指摘し、英語文章読解における受検者の誤答を把握することから始める必要性について述べた。2.2節では研究1として、過去に大学入試問題として出題された英語文章読解テストの項目を用いて、誤答選択枝に含まれる誤答の特徴を把握した。その結果、文章中に記述されていないことがらを述べている、文章中の記述に否定語や対義語を付している、文章中の内容や因果関係・比較対象を取り違えているなどの誤答を含んだ選択枝が作成されて

いたことが明らかとなった。2.3 節では研究 2 として、上記の誤答の特徴を含む錯乱枝を実験的に作成したテスト項目に対して、英語を第二言語とする日本の大学生から解答を得た。その結果、能力低群では英語文章中に記述されていないことがらを含む錯乱枝の選択率が高かったこと、能力中群では英語文章中の記述に否定語を追加したり因果関係を逆転させた錯乱枝の選択率が高かったこと、能力高群では英語文章中の記述に対義語を付した錯乱枝の選択率が高かったことが明らかとなった。2.4 節では、これらの知見が実際のテスト項目作成にすぐに活用可能な知見であることについて述べた。

第 3 章では、第 2 章の研究知見を精緻化し、選択枝の内容に関連する文章中の箇所（キーセンテンス）の特定の仕方が異なる設問の下で、錯乱枝の選択率を能力群別に検討する実験を行った。3.1 節では、研究 1・2 で残された課題として、錯乱枝の選択率が設問設定の影響を受ける可能性について指摘した上で、選択枝の内容を英語文章中で見つけ出すプロセスを考慮して設問を作り分けた際の錯乱枝の選択率を検討する必要性について述べた。3.2 節では研究 3 として、設問がどのような形で受検者にキーセンテンスの特定を求めているかについての調査を行い、「一内容（隣接する 2～3 文）のみで完結する設問」「一段落を踏まえる設問」「複数段落にまたがる設問」「文章全体を踏まえる設問」の 4 種類の設問があることが示された。3.3 節では研究 4 として、研究 3 での分類をもとに設問を作り分け、否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の選択率について、設問の種類間・能力群間での比較を行った。その結果、設問の種類間の違いはいずれの能力群においても否定語錯乱枝の選択率に影響を及ぼしていなかったこと、能力中群では複数段落にまたがる設問や文章全体を踏まえる設問で対義語を含む錯乱枝の選択率が高かったこと、能力高群では文章全体を踏まえる設問で因果関係の取り違いを含む錯乱枝の選択率が高かったことが明らかになった。3.4 節では研究 5 として、選択枝中に含まれる語がキーセンテンス中に含まれるかどうか、キーセンテンスの特定、ひいては選択枝の正誤判断に影響を及ぼす可能性を想定し、選択枝中にキーセンテンス中の語を含む重複条件と、その語を類義語に置き換えた非重複条件を設定して、否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の選択率を能力群別に検討した。その際、キーセンテンスと選択枝を見比べるという処理が、キーセンテンスの特定を明示的に求めるかどうかによって知見を踏まえて、キーセンテンスの特定を明示的に求める下位レベル設問と、キーセンテンスの特定を明示的には求めずに、段落の主旨や文章全体の構造の理解を求める上位レベル設問を作成した。その結果、錯乱枝の種類を問わず、能力低群では下位レベル設問で重複条件の錯乱枝の選択率が高く、上位レベル設問では非重複条件の錯乱枝の選択率が高いことが明らかになった。3.5 節では第 3 章のまとめとして、設問に応じて効果的となる錯乱枝を配置することの必要性についての議論を行った。

第 4 章では、英語文章読解能力の中でも段落や文章の要点把握の側面に着目し、英語文章を大域的に読解する際に生じる受検者の誤答の特徴を調査した上で、その特徴を含んだ

錯乱枝の選択率を能力群別に検討した。4.1節では、英語文章の要点把握に関する認知心理学・第二言語習得の知見が蓄積されているにもかかわらず、こうした能力を測るテスト項目についての教育測定学的検討がなされていない点を指摘し、英語文章の要点把握の側面を測るテスト項目の検討の必要性について述べた。4.2節では研究6として、Kintsch & van Dijk (1978) の文章要約モデルを参考に、段落および文章全体の要約を求める問題を作成し、英語を第二言語とする日本の大学生の解答から、英語文章の要約に関する以下の3つの誤答の特徴を明らかにした。第1に、段落中の重要でない部分を削除するプロセスに焦点を当てた問題では、必要な要素が不足している要約文が存在していた一方で、重要でない情報が混入した要約文も一定数存在していた。第2に、段落中の具体的な記述を一般化して要約するプロセスに焦点を当てた問題では、具体的な記述をそのまま抜き書きしている、適切でない一般化表現を含む等の特徴が、多くの受検者の要約文に見られた。第3に、段落中の記述内容を代表する文（トピック・センテンス）を産出するプロセスに焦点を当てた問題では、筆者の意図を一面的に記述している、筆者の意図から逸脱した記述を含むなどの特徴が受検者の要約文に見られた。この結果を踏まえ、4.3節では研究7として、上記の特徴を含んだ錯乱枝を実験的に作成し、能力群間で選択率の比較を行った。その結果、次の3点が明らかにされた。第1に、必要な要素が不足した要約文を錯乱枝として提示した場合に、能力群間での選択率の差が大きいこと、第2に、具体例やエピソードなどの記述をそのまま抜き書きした要約文を錯乱枝として提示した場合に、能力中群と能力高群との間で選択率の差が大きいこと、第3に、筆者の意図を一面的に記述していたり、筆者の意図から逸脱する内容が含まれる要約文を錯乱枝として提示した場合に、能力低群と能力中群との間で選択率の差が大きいことが示された。4.4節では、第4章のまとめとして上記3つの知見を整理した上で、大域的な側面の英語文章読解能力を測る際の項目形式として、受検者の誤答を踏まえた錯乱枝を含む多枝選択式項目のコストパフォーマンスの高さ等について議論した。

第5章では、上の7つの研究知見の整理を行い、これらの知見が大規模かつハイステークスなテストにおける項目作成を支援する役割を果たしうる点、受検者の誤答に関する調査に基づいて錯乱枝に含めるべき誤答を具体的に提示した点等から、学術的・実践的意義を有することについて述べた。本論文の研究知見は今後、認知診断を目的とするテスト項目作成に活用することも可能であるし、自動項目生成（Automatic Item Generation）の際の項目テンプレートの作成の際に活用することも可能である。受検者の誤答に焦点を当てた本論文の知見は、錯乱枝の効果的な作成に貢献することを通して、テスト得点に対して適切な解釈を行うことに寄与したものと考えられる。また、予備テストを行うことが困難なわが国において、テスト項目が統計的な基準を満たすための一つの視点を提供したものと考えられる。



# 第1章

## 序 論

本論文の目的は、多枝選択式の英語文章読解項目を題材に、受検者の典型的な誤りを反映した選択枝（錯乱枝）の作成手法の提案を目指して、錯乱枝の選択率を受検者の能力群別に検討することである。本章では、問題意識や理論的なレビュー・先行研究の概観などを行い、本論文の焦点を明確にしなが、各章の構成について述べる。

本章は、大きく4つの部分から構成される。1.1節「能力の測定におけるテスト項目の品質の担保」では、学力や能力を測定するテスト項目の品質について議論することの必要性、わが国におけるテスト項目の品質管理の現状、妥当性の高い測定を可能にするテスト項目の作成のために認知心理学の知見を利用することの有用性などについて概観し、テスト項目作成の原則を確認する。1.2節「多枝選択式項目の特徴と錯乱枝の機能」では、大規模かつハイスタークスのテストにおいて用いられることの多い多枝選択式テストの特徴について整理し、多枝選択式項目を用いた場合の錯乱枝の機能について、先行研究のレビューを通して整理する。1.3節「英語文章読解能力とその認知モデル」では、本論文が測定対象とする能力を明確にしなが、受検者がどのようにして英語文章を読み、項目に解答するののかに関する整理を行う。本章の最後に、1.4節「本論文の目的と構成」では、本章での整理を踏まえて本論文全体の目的を明確にした上で、本論文における構成と各章の概要について述べる。

### 1.1 能力の測定におけるテスト項目の品質の担保

一般に、何らかの能力を測定したいと考えたとき、多くの場合テストが用いられる。テストを用いて関心下の能力を測定するときには、複数のテスト項目を用意し、受検者から解答・応答を得ることとなる。解答・応答が得られたら、それらを事前に定めたルールに沿って得点化し、測定モデルを通じて関心のある能力についての推論を行うことになる。構成概念としての能力について適切な推測を行うためには、テスト項目や得点化のルール、測定モデルに関する検討を十分に行う必要がある。本節では、能力についての推測を

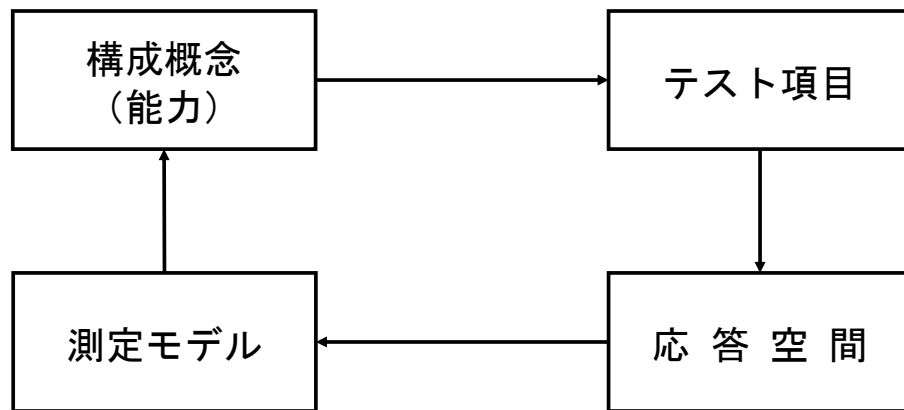


Figure 1.1. Wilson (2005) における項目応答の4要素モデル

行うための理論的なモデルについて紹介し、テスト項目の品質に関する議論の前提を確認する。

Wilson (2005) は、受検者の応答から能力に関する推測をする際のプロセスとして、項目応答の4要素モデル (four building blocks for item response) を提唱している (Figure 1.1)。このモデルでは、構成概念 (construct)・テスト項目 (item)・応答空間 (outcome space)・測定モデル (measurement model) の4つの要素から構成され、テスト項目・応答空間・測定モデルを通じて能力に関する推測を行うプロセスが想定されている。

モデル中の「構成概念 (能力)」では、関心下の能力そのものに関する詳細な検討が行われる。能力に関する検討の具体的な例としては、その能力に関連する理論のレビューを行う、類似した能力を測定した事例を確認するなどが挙げられる。このような検討の後、受検者にできること・できないことを具体化し、能力水準を表す連続的な尺度の上に位置づけた構成概念マップ (construct map) を作成することとなる。

「テスト項目」の段階では、構成概念マップに即して具体的なテスト項目を作成する作業が行われる。意図した応答を最大限引き出すことのできるテスト項目を作成するため、テスト形式についての検討や受検者の認知プロセスの検討などが行われる。これらの情報を得るために、受検者への調査やインタビュー、発話思考などが一般的に行われ、場合によっては反応時間の測定や視線の測定 (e.g., Bax, 2013) などの方法が用いられることもある。さまざまな方法を通じて、受検者のもつ能力とテスト項目に対する応答との間の因果関係を構築する。

「応答空間」では、テスト項目に対して受検者から得られた応答を、構成概念としての能力に関する推測に使用できるように処理を行う。受検者が選んだ選択枝の番号そのもの



や、記述式項目に対する解答データなど、項目単位の生の応答をそのまま構成概念に対する推測に用いることは難しいため、事前に定めた得点化のルール・採点基準に沿って項目得点を算出し、構成概念ごとにまとめ上げるなどして、構成概念としての能力に対する推測の準備を行う必要がある。

最後に「測定モデル」では、受検者の得点と構成概念としての能力を結びつけ、「構成概念（能力）」に関する情報を得る。測定モデルとしてよく知られているものに、古典的テスト理論における基本式（テスト得点  $X =$  真の得点  $T +$  誤差  $E$ ）やラッシュモデル・項目応答モデルなどが挙げられる。特に、ラッシュモデルや項目応答モデルにおける受検者の潜在特性値  $\theta$  は、構成概念としての能力を定量的に表現し、能力に対する推論を強力に支えるもののひとつである (Wilson, 2005)。

能力に対する推論が適切であることを示すためには、Wilson (2005) における「構成概念」「テスト項目」「応答空間」「測定モデル」のすべてに対する検討が重要であり、これら4つの要素が密接に結びついていることを示す必要がある。一方で、わが国における大規模試験の多くは、このようなモデルに基づく能力測定を行う制度設計にはなっていない。モデルを仮定せずに能力測定に関する議論を行った場合、しばしば論点が曖昧化しがちである一方で、Wilson (2005) のようなモデルに沿った議論を行うことにより、モデル中の要素のうち検討が十分でない箇所を特定でき、論点の明確化につながるものと考えられる。このような点において、Wilson (2005) のようなモデルに基づいて能力測定に関する議論を行うことは極めて重要であると言える。

本論文では、モデル中の4要素のうち「テスト項目」に焦点を当てる。Wilson (2005) のモデルにおいて、テスト項目は「能力」と「応答空間」の間をつなぐ役割を担っている。すなわち、ターゲットとする能力を定量化する際の媒体として、テスト項目は重要な役割を担っていると言える。テスト項目の作成について検討を行うことは、関心下の能力の高低と可能な限り対応する得点を得ることについて検討を行うことであると考えられる。

Wilson (2005) のモデルでは、テスト項目が測定したい能力を反映するように作成するということが想定されている。上でも述べたように、測定したい能力と受検者の応答との因果関係が構築されるようなテスト項目を作成することが必要となる。加えて、Wilson (2005) のモデルが一連のサイクルになっている点からは、応答空間の得点を用い、測定モデルに基づく能力についての推論を行った結果を踏まえてテスト項目を改善するという示唆される。テスト項目の作成にあたり、能力と受検者の応答との密接な関係を構築するために、事前に受検者の応答を得ておく必要もあるということになる。この点は、テスト項目の品質をどのように捉えるかという点と関連する。本節では、測定したい能力に対応したテスト項目を作成するときの重要な検討事項としてテストの品質を取り上げ、教育測定学における理論と国内外の現状の整理を行うこととする。

### 1.1.1 テスト項目の品質

教育測定学では、テスト項目の品質として、定量的な側面と定性的な側面の2つの観点があると言われている(加藤, 2014)。定量的な側面からみたテスト項目の品質とは、困難度や識別力、信頼性、測定精度などの統計的性質のことを指し、定性的な側面からみたテスト項目の品質とは、テスト項目が問う内容(内容的妥当性)やテスト得点の解釈、教育的・社会的影響(結果妥当性)など、妥当性の種々の側面を指す。テスト項目の品質を検討する際、定量的な側面と定性的な側面のどちらに焦点を当てるかということによって、テスト項目に対するアプローチの仕方は大きく異なる。

定量的な側面からテスト項目の品質を検討する場合、実際の受検者集団に近い受検者に対して予備テストを実施し、得られた解答データを用いて項目分析(item analysis)や信頼性・測定精度に関する確認を行うことが多い。項目分析では、困難度や識別力・信頼性などの指標が基準を満たしているかどうかについて判断を行うことになる。テスト項目が満たすべき基準の例としては、正答率が極端に高かったり低かったりしていないこと、識別力が一定程度高いこと、信頼性を著しく低めていないことなどが挙げられる。通常、上記の基準を満たした項目は、困難度・識別力・信頼性(項目情報量)などの情報とともに、大量のテスト項目が蓄積された項目バンクに追加され、実際のテストで使用される項目となる。それに対し、基準を満たさなかった項目には、設問や選択枝などの修正・変更を行い、再度予備テストを実施して基準を満たしているかどうか、判断を行うことになる。

一方で、定性的な側面からテスト項目の品質を検討する場合には、テスト項目の内容的妥当性(e.g., Cronbach, 1984; Messick, 1995)や結果妥当性(Messick, 1995)にアプローチすることとなる(加藤, 2014)。このようなアプローチを通じてテスト項目の品質の高さを示すためには、テスト項目から算出される得点の解釈と利用(interpretation and use)が妥当であることを確認する必要がある(Kane, 2013)。具体的には、受検者の実際の学習場面に関する調査、テスト項目の出題内容や出題範囲の適切な設定、受検者の解答プロセスに関する検討を行うことなどが考えられる。さらに、テストがこれらの点を満たしているかどうか確認するために、テスト得点の解釈と利用について論証すべき命題(interpretation/use arguments; IUA)を作成することなどが挙げられる(Kane, 2013)。出題内容や出題範囲、受検者の解答プロセスについて検討を行うことは、Messick(1995)の妥当性の諸側面のうちの内容的側面や実質的側面に着目することを意味し、テスト得点の使用目的などについて検討を行うことは、Messick(1995)の一般化の側面や結果の側面に着目することを意味する。

一般に、テスト開発の段階では受検者の解答データが得られていないため、定性的な側面からテスト項目の品質を担保する手続きがとられた後、予備テストを実施し、その

後の段階では定量的な側面からテスト項目の品質を確かめる手続きがとられる。いわば、Wilson (2005) において提示されているサイクルを本番の前に一度回し、能力に対する推論を適切に行うことができるかどうか確認していると言えるだろう。定性・定量の両方の観点からテスト項目に対する品質管理を行うことは、能力に対する推論の適切さを高める重要な手続きであると言える。

### 1.1.2 わが国におけるテスト項目の品質管理の現状

前節で述べたテスト項目の品質管理の手法は、主に海外の大規模テストにおいて用いられてきた。一方で、わが国ではテスト項目の品質管理を十分に行うことが難しい状況も多く存在する。その背後には、わが国のテストをめぐる文化や信念（テスト文化）の影響が想定される。そこで本節では、わが国のテスト文化を概観しながら、テスト項目の品質管理の現状について整理する。

わが国においてテスト項目の品質管理を困難にする 1 つ目の要因は、テスト実施後にテスト項目を公開することが望ましいとされていることである (石井, 2018; Kuramoto & Koizumi, 2016)。米国をはじめとする海外の大規模テストでは、テスト実施後にテスト冊子を回収しているため、露出するテスト項目の数を最小限に抑えることが可能である。このような状況下では、予備テストで基準を満たしていることが確認されたテスト項目を本番のテストでも使用できる。一方、わが国ではテスト項目を公開することが望ましいとされ、テスト実施後には原則としてテスト項目が公開される。例えば、大学入試センター試験のテスト冊子は受検者本人の持ち帰りが可能であり、翌日の新聞にはテスト項目が掲載される (石井, 2018)。このように、テスト項目の公開に対する要求が高いわが国では、仮に予備テストを実施した場合、教育産業などによってテスト項目の復元が容易に行われるため、受検者自身の練習・学習などが想定され、最終的に項目の統計的性質が予備テストの時点のものとは異なる可能性も考えられる。こうしたわが国の状況は予備テストを困難にするため、受検者の解答データを事前に得ることができず、テスト項目の品質を定量的な観点から検討することができない状況となりがちである。ただしわが国においても、テストの仕様としてテスト項目を非公開としている試験も存在する (e.g., 医療系大学間共用試験; 石井, 2018) 点を踏まえると、わが国においてテスト項目の公開に対する要求を和らげることも可能であると考えられる。

わが国のテストにおける項目の品質管理を阻害する 2 つ目の要因は、項目作成を主たる職業とする者 (item writer) の不在が挙げられる。米国など海外の大規模テストでは通常、測定内容そのものの専門家であり、かつ項目作成等に関する十分なトレーニングを受けた、項目作成を専業とする者がテスト項目を作成する (Downing, 2006b)。一方、わが国の大規模テストでは、他の業務も併せもつ者がテスト項目を作成することが多い。そのた

め、トレーニングが十分でない状態でテスト項目を作成することになり、テスト得点の解釈や結果の利用の適切さなど、定性的な側面からみた品質を意識しないテスト項目が作成されうる状況となる。もちろん、作成されたテスト項目の中には、これらの点を十分意識して作成されたものも存在するが、多くの場合、テスト項目の定性的な側面について精査するような機会は少ないと考えられる。このような点を踏まえると、わが国においてテスト項目の品質を定性的な側面から管理することも十分でない状況であると考えられる。

その他にも、テスト項目の品質に影響を及ぼすわが国の特殊な状況は存在するが、上記の2点のみを取り上げても、海外における大規模テストの状況とは大きく異なっており、これらがテスト項目の品質管理に大きな影響を及ぼしていることが見て取れる。わが国でテスト項目の品質管理を行うためには、次の2つの解決策が考えられる。第1の解決策は、定量的に品質管理ができるよう、テストをめぐる状況そのものを変えることである。ただし、社会や文化そのものの変革が必要であり、テスト実施上の制約との兼ね合い等から、そのような解決策を採ることが厳しい場面もありうる。第2の解決策は、定性的な側面からみたテスト項目の品質管理の状況を改善することである。例えば、項目作成者に対するトレーニングの充実や受検者の解答プロセスを踏まえた項目作成手法の提案、テスト得点の利用方法を見据えたテスト開発などが挙げられる。第2の解決策は、項目作成の段階に関する手続きであるため、時間の制約が厳しいものでなければ実施可能である。次節では、定性的な観点からのテスト項目の品質担保に焦点を当て、妥当性の証拠を得ることのできるテスト開発の条件について整理する。

### 1.1.3 妥当性の証拠を得ることのできるテスト開発

定性的な観点からテスト項目の品質を担保しようとするとき、上述のように、内容的妥当性や結果妥当性の側面から検討を行うこととなる。このような検討は、実際に項目作成を行う前の段階であるテストの全体計画 (overall plan) や出題内容の決定 (content definition)、テストの仕様の決定 (test specification) の段階で行われるものである (Downing, 2006b)。これらの項目作成手続きにおいて、妥当性の証拠 (validity evidence) が得られるようにテストを設計することにより、テスト項目の品質が担保され、構成概念 (能力) と応答空間 (Wilson, 2005) の結びつきが高まる。本節では、妥当性理論の変遷を紹介しながら、妥当性の証拠を得てテストの品質管理を行っている事例を確認し、テスト項目の作成の段階で得るべき知見についての示唆を得る。

そもそも妥当性とは、測りたいものをきちんと測れているかということを目指す (石井, 2014)。妥当性を確認するにあたっては、測りたいものが測れているという状況証拠 (validity evidence) を多様な観点から収集することが必要である。これまでに、多くの計量心理学者によってさまざまな観点が提示され、測りたい能力がきちんと測定されているこ

とを示すための枠組みが精緻化されてきた。

教育測定学・計量心理学の領域で一般的に知られているのは、Cronbach & Meehl (1955) の妥当性理論である。Cronbach & Meehl (1955) は、予測的妥当性 (predictive validity) ・併存的妥当性 (concurrent validity) ・内容的妥当性 (content validity) ・構成概念妥当性 (construct validity) の4側面から妥当性を捉える枠組みを提唱した。このうち内容的妥当性は、テストの出題内容からの適切なサンプリング (領域代表性) について証拠を収集する点で、テストの内容に関わる妥当性であると言える。一方、測定したい能力に関する受検者の認知過程の確認や、テスト結果の利用に関する証拠の収集に関しては、Cronbach & Meehl (1955) において言及されておらず、単に測定したい能力に関する理論の確認について言及されるのみにとどまっている。

その後、Whitely (1983) は、心理学において情報処理のプロセスに関する知見が集積されてきたことを取り上げ、心理学の理論に合わせて Cronbach & Meehl (1955) の妥当性を変更することの必要性を指摘している。具体的には、心理学の理論が原因—結果の関係を分析する視点から、人間の認知の背後にあるプロセスを推論して行動を説明する視点に移り変わったことを踏まえ、何らかの課題 (テスト項目) に取り組む際のプロセスやメカニズムなどが、その課題に反映されているかどうかを検討対象とする必要があると述べている (Whitely, 1983)。Whitely (Embretson) 自身は、この観点を「構成概念の表現 (construct representation)」と呼んでおり、想定される認知プロセスが課題 (テスト項目) のパフォーマンスを大きく左右することを確認することも妥当性の証拠のひとつであると主張している (Whitely, 1983)。Whitely (Embretson) は、Cronbach の内容的妥当性に情報処理プロセスという視点を追加したとともに、妥当性を課題 (テスト項目) ごとに検討することの必要性を提言している点で、妥当性理論に大きなインパクトをもたらしたと言えるだろう。なお、Cronbach は後に、構成概念妥当性の検討方法のひとつとして、発話思考法 (think aloud) を用いた誤差の低減に言及している (Cronbach, 1984)。

Whitely (Embretson) の指摘から程なくして、Messick (1995) が妥当性の6側面を提唱している。具体的には、測定内容の関連性や領域代表性、項目形式や採点基準などの測定論上の質に焦点を当てた「内容的側面」、テスト項目を解答する際の認知過程に焦点を当てた「実質的側面」、得られた得点間の構造に焦点を当てた「構造的側面」、母集団やテストの状況、課題の違いによらず同一の結果が得られるかどうかに焦点を当てた「一般化の側面」、収束的妥当性や弁別的妥当性に焦点を当てた「外的側面」、テスト得点の解釈と利用に焦点を当てた「結果的側面」である。これらのうち、受検者の認知過程を取り上げた実質的側面や、テスト得点の利用を取り上げた結果的側面などが登場しているのが、Messick の妥当性理論の大きな特徴である。

Kane (2013) の妥当性は、「テスト得点の解釈と利用 (the interpretation and use of test scores)」に特徴づけられる。テスト得点の解釈にあたっては、テスト得点の解釈と利用に

Table 1.1  
TOEFL iBT における妥当性の命題と対応する証拠

命題	妥当性の証拠
<ul style="list-style-type: none"> <li>● テストの内容は、受検者が大学の学習の中で出会うさまざまな課題・文書・話し言葉と関連するものであり、かつこれらを代表するものである</li> <li>● テストにおける課題や採点基準は、受検者の学術的な言語能力を証拠づけるものとして適切である</li> <li>● テストの課題への応答時にみられる言語知識や言語の処理過程・方略の使用などから、受検者の学術的な言語能力の様相がわかる</li> <li>● テストの因子構造は、英語技能間の関係に関する理論上の見解と一致する</li> <li>● テストにおけるパフォーマンスは、学術的な言語能力に関する他の指標・基準との関連がある</li> <li>● テストの結果は適切に利用され、よい帰結をもたらすものである</li> </ul>	<ul style="list-style-type: none"> <li>● 英語を用いて講義や演習が行われている大学での言語使用に関する先行研究のレビュー</li> <li>● 課題やテストのデザインに関する予備的研究 記述式答案や発話に対する採点基準の検討</li> <li>● 文章読解問題に解答しているときの受検者の記述式答案や発話の特徴に関する検討 解答の際に使用している方略についての検討</li> <li>● 予備テストの受検者の応答に対する因子分析</li> <li>● 自己評価の結果、大学でのプレースメント・テストをはじめとする外在基準とテスト得点との関連の検討</li> <li>● テストに向けた準備や、テスト得点に対する適切な解釈を補助する課題の開発 テストの影響力に関する長期的な実証研究</li> </ul>

ついて論証すべき命題 (IUA) を作成し、それらを論証できるように受検者から解答データを収集し、証拠を揃えることが必要であると述べている (Kane, 2013). IUA の例としては、受検者から観測されたパフォーマンスを得点に変換するための「得点化・採点に関する命題」、テスト項目や課題場面を超えて受検者のパフォーマンスを表すための「得点の一般化に関する命題」などが挙げられる。また、テスト得点の利用に関しては、事前に想定されていた用途として得点を使用されているかどうかの評価が必要であるとされる (Kane, 2013).

妥当性の証拠収集を行っているテストの一例として、TOEFL iBT が挙げられる。TOEFL iBT は、L2 学習者が北米を中心とする英語圏の大学・大学院への入学を希望する場合に、自らの英語能力の程度を示すことを目的として利用される。TOEFL iBT の開発にあたっては、上記に示したテストの目的に照らして 6 つの命題を設定し、その命題を支持する証拠として示すべきことがらとの対応が示されている (Educational Testing Service, 2008). TOEFL iBT では、これらの命題を満たすことができるようなテストの設計および事前の検討・分析が行われている。Table 1.1 に、妥当性に関する命題と対応する証拠の一覧を示した。

Table 1.1 に挙げられている 6 つの命題について整理する。1 番目の命題は、測定内容の

領域代表性や課題の真正性に関する命題である。L2 学習者が北米の大学で学ぶ中で必要な英語能力を特定し洗練していく作業であり、テストの解釈および利用の両方を見据え、課題の真正性を担保するために設定されている (Educational Testing Service, 2008)。2 番目の命題は、受検者のテスト得点を生み出す課題および採点基準に関するものであり、受検者のパフォーマンスに対して適切な得点を与えるかどうかを検討対象としている。3 番目の命題は、Whitely (1983) や Messick (1995) の強調点である受検者の認知プロセスを、4 番目の命題は Cronbach & Meehl (1955) の構成概念の一側面として挙げられている因子的妥当性を検討対象としており、いずれもテスト得点の解釈の根拠を理論的に担保するために設定されている (Educational Testing Service, 2008)。5 番目の命題は 1 番目の命題と同様、テストの解釈および利用の両方をターゲットとした命題であり、Cronbach & Meehl (1955) の基準関連妥当性の側面に焦点を当てている。6 番目には、テスト得点の利用の側面を見据えた命題が設定されている。

TOEFL iBT の例はあくまで一例にすぎず、テストの仕様や利用目的に応じて収集すべき妥当性の証拠は異なる (Kane, 2013) が、TOEFL iBT における妥当性の証拠収集の観点からは、他のテストを開発する際の参考資料として機能するものと考えられる。上記に示したテストの妥当性に関する証拠の収集を総合的に行うためには、膨大な研究知見の蓄積が必要である。したがって、妥当性に関する証拠の収集を総合的に行おうとすると、本論文で扱うことのできる範囲を大幅に超えることが考えられる。そこで本論文では、わが国のテスト項目の品質管理に関する現状を踏まえ、検討対象とする妥当性の側面を絞り、妥当性の証拠に資する実証的知見の一部を提供することをねらいとする。

上に見たように、予備テストを実施することが困難なわが国においては、テストを実施する前に予備テスト等を通じて受検者の解答データを得ることが困難であるため、現状では Table 1.1 に示した 4~6 番目の命題に対する妥当性の証拠を得ることが難しい状況にある。一方で、測定対象とする能力に関連する先行研究のレビューを行い、受検者の解答プロセスの把握などに基づいた項目作成を提案できる研究知見を蓄積すること自体は、わが国においても一定程度可能であると考えられる。本論文では、わが国の現状を考慮した場合に得ることが可能な妥当性の証拠のひとつとして、Table 1.1 の 3 番目の命題として挙げられているような、受検者の認知・解答プロセスに焦点を当てたテスト項目の作成に焦点を当てる。この側面の妥当性について検討を行うことにより、テスト項目の定性的な品質を担保することに寄与するものと考えられる。

#### 1.1.4 項目作成における認知心理学の役割と本論文の立ち位置

前節ではテスト開発や項目作成の段階において、測定対象の能力に関連した先行研究のレビューや受検者の解答プロセスの把握を行うことが、妥当性を証拠づけるもののひとつ

となる可能性について述べた。このような考え方にに基づき、測定対象の能力に関する認知心理学の知見を利用した項目作成が、2000年代から注目されてきた (e.g., Embretson, 2001; Gorin, 2005; Gierl & Haladyna, 2013; Haladyna & Rodriguez, 2013)。上の先行研究の多くは、読解やライティング、数学的能力、問題解決能力などの認知的能力を測定する場合に、認知心理学に基づいた項目作成を行うことで妥当性を高めうる可能性について示唆している。このように、認知心理学に基づいてテスト項目の仕様について検討を行う考え方は、認知心理学の原則 (cognitive psychology principle) あるいは認知デザインシステム (cognitive design systems) と呼ばれる (Embretson, 2001)。本節では、項目作成における認知心理学の原則の主要な点について述べた上で、このような項目作成の手法の利点について整理する。

認知心理学の原則の主要な点の1つ目は、測定対象の能力に関する検討を項目作成の前に実施することである。これまでのテストにおいて、測定対象の能力に関する検討にあたっては、受検者の解答データを用いて法則定立的な距離 (nomothetic span) を確認するという方法が取られてきた (Embretson, 2001; Whitely, 1983)。法則定立的ネットワークを検証する目的でよく用いられるのが、因子分析である。このような方法で測定対象の能力に関する検討を事後的に行った場合、往々にして項目作成時に想定していた因子構造と異なることがある。このような状況ではこれまで、項目の統計的指標 (e.g., 識別力の高さ) や信頼性が優先されてきたが、そうした対応をとることでテスト得点に対して当初とは異なる解釈が付与される可能性があり、妥当性を著しく損なうこともある (e.g., Embretson, 2001)。以上のように、従来のアプローチでは、テスト実施後でないと測定対象とする能力についての理論的根拠を確認することができないという問題点を抱えていると言える。

一方、認知心理学の原則に基づいて項目作成を行う場合、テストの仕様を決定する段階において認知心理学の知見を利用し、測定対象とする能力を理論的に定義した上で項目作成を行うこととなる。この方法を適用すれば、項目作成の段階で受検者の認知処理に関する見通しが立ち、テスト得点に対して項目作成時とデータ収集後で一貫した解釈が可能となる。妥当性の重要な側面として主張されているテスト得点の解釈 (Kane, 2013) をより適切なものとする点において、認知心理学に根差した項目作成の利点は大きいものと考えられる。

認知心理学の原則の主要な点の2つ目は、受検者の認知処理を反映したテスト項目の作成が可能であることである。テスト項目を作成するときには、内容そのもの (content) と認知的要求度 (cognitive demand) の双方に関する検討が必要である (Haladyna & Rodriguez, 2013)。認知的要求度とは、典型的な受検者がそのテスト項目に取り組む際の認知処理の複雑さ (mental complexity) と定義される (Haladyna & Rodriguez, 2013)。テスト項目の内容については、教科の専門家の判断が重要とされる一方で、認知的要求度については、テスト開発の段階で受検者の認知処理を理解する必要が生じる。受検者の認知処理を理解する



手助けとなるのが、認知心理学のモデルである。認知心理学の一部の領域ではこれまで、読解やライティング（作文）、数学的能力や問題解決能力などの背後にあるメカニズムが解明されてきた (Haladyna & Rodriguez, 2013)。認知心理学の原則に基づく項目作成の枠組みでは、認知心理学における認知モデルや情報処理モデルが、それぞれのテスト項目に対する正誤を一定程度説明するという発想の下で、項目作成を行うことになる。教科の専門家の観点だけでなく、受検者がテスト項目に解答するときの認知処理という観点をそれぞれのテスト項目に反映させることにより、認知処理のうちでどの段階までを流暢に行うことができたかということが如実にテスト得点に反映され、テスト得点の解釈をより適切なものにするに寄与することが期待される。

以上の点を踏まえると、認知心理学の原則に基づいた項目作成は、テスト得点の適切な解釈に貢献し、定性的な側面での高い品質を担保できる方法であると考えられる。このような方法で作成されたテスト項目は、受検者個人の得手不得手を診断する認知診断 (cognitive diagnosis) や自動項目生成 (automatic item generation) に役立てることが可能であるとの見地から、近年の教育測定学研究において注目を集めている (Embretson, 2001; Gierl & Haladyna, 2013; Haladyna & Rodriguez, 2013)。上記のような研究背景を踏まえると、教育測定学における項目作成の技術について検討を行うことは、品質の高いテスト項目を用いた測定を可能にする点で、大きな意義があるものと考えられる。

本論文では、定性的な側面からみたテスト項目の品質の担保を目指し、認知心理学の知見に基づいて項目作成を行うという立場をとる。具体的には、項目作成の前の段階で測定対象の能力に関する認知過程や情報処理過程に関するレビューを行うという点、受検者の認知処理をテスト項目に反映させるという点を重視して、項目作成に関する検討を行う。なお、本論文で取り上げる英語文章読解能力に関連する認知心理学の知見のレビューは、1.3 節で行う。

1.2 節では、大規模試験で頻繁に用いられるテスト形式である多枝選択式項目について整理する。多枝選択式項目を作成する場合、選択枝の作成、とりわけ誤答選択枝（錯乱枝）の作成に多くの時間と労力が費やされる (Haladyna & Rodriguez, 2013)。この状況を解決する方法として、本節で述べた認知心理学の原則を適用する。具体的には、項目作成の前に測定対象とする能力のレビューを行い、認知過程や情報処理過程を経る中での受検者の誤りを誤答選択枝（錯乱枝）に反映させる手続きをとる。この手続きの詳細については、1.2.4 節で述べる。

## 1.2 多枝選択式項目の特徴と錯乱枝の機能

わが国に限らず、膨大な数の受検者がテストを受検するような大規模試験で、かつ個人の処遇を左右するハイスタークスのテストでは、多枝選択式テストがよく用いられる。わ

が国における大規模かつハイステークスなテストの例として、大学入試センター試験が挙げられる。大学入試センター試験はおよそ 58 万人の受検者数を抱える大規模試験である(大学入試センター, 2018)。大学入試センター試験では、テスト得点そのまま志望大学の合否に直結するハイステークスな試験であるという点などから、採点の客観性やスピードが求められるため、多枝選択式テストが用いられている。

大規模かつハイステークスなテスト以外でも、教育現場において多枝選択式テストが用いられることがある。例えば、各学校で実施される定期テストは、特定の学校内における学年・学級単位で実施される小規模なテストであるが、成績評価の判断材料として定期テストのスコア以外の成果物なども含まれる状況であるため、ハイステークスとは言い難い。このようなテストにおいても、多枝選択式項目が用いられることがある。

以上に見るように、多枝選択式テストはさまざまな教育場面で用いられており、現代に生きる私たちにとってなじみ深いテスト形式であると言える。しかし、多枝選択式問題の作成の際には、多くの点に留意する必要がある(Haladyna et al., 2002; Haladyna & Rodriguez, 2013)。多枝選択式テストを用いて何らかの能力を測るにあたっては、多枝選択式項目そのものについての基礎的・本質的理解を深める必要がある。そこで本節では、1.2.1 節で多枝選択式項目そのものがもつ特徴を整理し、1.2.2 節で多枝選択式項目の長所と短所を理解しながら、1.2.3 節・1.2.4 節で選択枝の役割、なかでも魅力的な誤答選択枝として位置づけられる錯乱枝(distractor)の機能について概観する。

### 1.2.1 多枝選択式項目の特徴

多枝選択式項目は、設問幹(stem)と複数の選択枝(options)から構成されている。設問幹とは、質問文の形式、あるいは不完全文の形式で、受検者に求めることがらを指定する部分のことである(Haladyna & Rodriguez, 2013)。選択枝とは、設問幹で示されている質問文に答える文、あるいは設問として提示された不完全な文を補う文のことである。複数の選択枝のうち、正しい選択枝・最も適切な選択枝は正答選択枝(correct option, key)、誤りを含む選択枝は誤答選択枝(incorrect options, fuels, alternatives, distractors)と呼ばれている(Haladyna & Rodriguez, 2013; 肥田野, 1972)。正答選択枝は、その正しさが疑いなく唯一に定まる選択枝であることが求められ、誤答選択枝は、その誤りが明確な選択枝であることが求められる。

上記に示した特徴をもつ多枝選択式項目は多くのテストで頻繁に用いられるが、このようなタイプの項目形式は従来型の多枝選択式項目(conventional multiple-choice item)と呼ばれ、解答選択形式(selected-response format)のひとつに含まれる形式である。解答選択形式に属する他の項目形式として、真偽式(true-false format)、組み合わせ式(matching format)などが挙げられる。

本論文では、従来型の多枝選択式項目がよく用いられる項目形式であることや、従来型の多枝選択式項目を材料とした先行研究の知見が一定程度蓄積されていることなどから、設問幹・正答選択枝1つ・誤答選択枝複数から構成される、従来型の多枝選択式項目に焦点を当てることとした。なお、以降の記述において特に断りのない場合には、「多枝選択式項目」は従来型の多枝選択式項目を指すものとする。

### 1.2.2 多枝選択式項目の長所と短所

多枝選択式項目の長所として、1) 採点の客観性・容易性、2) 受検者の解答時間の短さ、3) 項目の領域代表性の担保が挙げられる。以下では、これら3つの長所について詳しく述べることにする。

第1に、多枝選択式項目の採点は客観的かつ容易である。多枝選択式項目は、解答構築式項目 (constructed-response items) に対する主観的・恣意的採点を解決するという経緯から登場したテスト形式である(池田, 2007)。項目作成の段階であらかじめ正答選択枝の番号を決めておくことにより、あらかじめ決めた正答の番号を選んだ受検者には得点を与え、その番号を選んでいる受検者には得点を与えないという得点化のルールを明確に定めることができるため、採点の主観性を排除することが可能となる。

また、多枝選択式項目の採点にあたっては、受検者の応答があらかじめ決めた正答選択枝の番号と一致するかどうかを確認すればよいから、採点者の時間や労力を割かずコンピュータでの採点を行うことが可能であるという点で、採点は非常に容易である。現在では、コンピュータの性能が非常に優れているため、採点のスピードはかつてと比べても格段に速い。このように、多枝選択式項目の採点の客観性と容易さは、大規模試験において、主観性を排除しながら迅速に採点を行うためのテスト形式として非常に有用である点から、今日の教育測定においても用いられるテスト形式となっている。

第2に、受検者の1項目あたりの解答時間が短くて済む点が挙げられる。用語や記号等の解答を求める短答式以外の解答構築式項目では、受検者に大きな課題を提示し、その課題を解決するプロセスを記述させることが必要となる。このため、解答構築式項目を用いる際には、多枝選択式項目を用いる場合よりも解答時間を長く設定し、受検者から十分に練られた解答を得ることが必要となる(池田, 2007)。一方で、多枝選択式項目の解答にあたっては、正答だと思う選択枝を選んでマークすることのみがアウトプットとして求められるため、1項目あたりの解答時間が非常に短くて済む。このような利点から、広い範囲にわたる多くの項目を受検者に提示することが可能となるため、信頼性の向上に寄与する。

第3に、測定したい能力の範囲を一定程度カバーできるように項目を提示できるため、領域代表性の担保につながる点が挙げられる。上に述べたように、多枝選択式項目を用い

ることで多くの項目を受検者に提示することが可能になるため、広い範囲にわたる出題領域からの項目をテストに収録することができる。テスト項目1つ1つは、測定したい能力全体のうちの限定された一部分しかカバーできないが、カバーする範囲の異なる項目が多く集まれば、全体として広い範囲をカバーしたテストとなる。多枝選択式項目を用いることによるこのような利点は、測定したい能力を的確に捉えることにつながり、構成概念妥当性の向上につながるものである。

多枝選択式項目を用いることのその他の利点としては、コンピュータを用いたテストを行う際の項目バンクの構築にあたり、低コストで項目が再利用可能であること、認知診断を行うための再得点化が可能であることなどが挙げられている (Haladyna & Downing, 1989)。以上に示した多くの点から、多枝選択式項目は、テスト理論の面においても、実用の面においても、多くの利点をもっていることがわかる。

一方で、多枝選択式項目の短所は、項目の作成が難しいという点である (Downing, 2006a; Haladyna et al., 2002; Haladyna & Rodriguez, 2013; 肥田野, 1972)。Downing (2006a) や Haladyna et al. (2002), Haladyna & Rodriguez (2013) は、多枝選択式項目の作成が非常に労力と時間を必要とする作業であることを指摘している。米国には、テストの項目作成を専業とする者（項目作成者）が存在するが、熟練した項目作成者であっても多大な時間と労力を要することが知られている (Downing, 2006a)。項目作成が非常に難しい作業である点を踏まえ、海外における教育測定学の先行研究では、実証的知見をまとめた項目作成ガイドラインが提示されているほどである (e.g., Downing, 2006a; Ebel & Frisbie, 1991; Haladyna & Downing, 1989; Haladyna et al., 2002; Haladyna & Rodriguez, 2013)。

上に示した多枝選択式項目の利点の多くは、項目作成をいかに丁寧に行うかということにかかっていると考えられる。実際、第1の利点として挙げた採点の客観性は、正答選択枝の出来次第では担保されないものになる可能性があるし、第2の利点に挙げた受検者の解答時間の短さについても、受検者にとってわかりにくい項目を提示した場合には、受検者に不必要な認知負荷をかけることになるため、不出来な項目への解答に余計な時間がかかることになりかねない。このようにして見ると、多枝選択式項目の長所が長所となりうるためには、十分に推敲・吟味された項目が必要であると言えるだろう。

以上を踏まえると、多枝選択式項目がその長所を十分に発揮するためには、項目作成の段階において、項目に対する十分な検討や吟味、推敲が必要であると考えられる。

### 1.2.3 多枝選択式項目における選択枝の役割

このように、多枝選択式項目を用いるにあたっては、項目作成の段階が最も重要なプロセスとなる。特に選択枝の作成は、非常に難しいとされている (Haladyna & Downing, 1989; Haladyna et al., 2002; Haladyna & Rodriguez, 2013; 肥田野, 1972)。例えば以下でも述

べるように、正答選択枝は教科の専門家の全員が正しいと認めるものでなければならない (Haladyna & Rodriguez, 2013). ただし、歴史のテスト項目の作成における教科の専門家間での歴史認識の相違や、科学のテスト項目作成における既存の知識体系の更新、文章読解項目の作成における筆者の意図の解釈の違いなど、疑いの余地なく正しい選択枝を作成することには労力がかかることもある。また、誤答選択枝は能力の低い受検者にとって正しく見え、能力の高い受検者には選択枝に反映された誤りがわかるように作成する必要がある (Haladyna & Rodriguez, 2013). 能力の低い受検者に選ばれる誤答選択枝を作成するためには、受検者がどのような誤信念や誤解をもつかについて理解する必要がある。受検者の誤信念や誤解を事前に想像することには限界があり、項目作成者が想定した誤信念や誤解が実際の受検者には生じないこともある。したがって、受検者の誤信念や誤解を反映した誤答選択枝の作成も非常に労力がかかる。

本節では、選択枝が上記のような特徴をもたなければならない理由を整理し、多枝選択式項目を用いた能力の測定において選択枝が果たす役割について概観する。

第1に、上述したとおり、正答選択枝は教科の専門家全員によって正しいと判断されるものである必要がある (Haladyna & Rodriguez, 2013). この点は、どのような能力を測定するテストの項目においても正答選択枝が満たす必要のある条件である。その理由として、教科の専門家によって正しいと判断された正答選択枝が、多枝選択式項目の利点の一つである採点の客観性を成り立たせる点が挙げられる。正答選択枝が教科の専門家から正しいと判断されなかった場合、あらかじめ定めた正答選択枝の番号を選んだ受検者にのみ得点を与えるという客観的なルールを適用することが難しくなる。多枝選択式項目においてあらかじめ定められた明確な得点化のルールは、正しいことが保証された正答選択枝の存在に依拠していると言える。このような点から、教科の専門家によって正しいと判断された正答選択枝は、多枝選択式項目の利点のひとつである採点の客観性を成り立たせることに寄与する。

第2に、多枝選択式テストにおける誤答選択枝は、受検者の典型的な誤りを反映したものであり、かつ能力水準の低い受検者にとって魅力的に見える必要がある (Haladyna & Rodriguez, 2013; Rodriguez et al., 2014). 誤答選択枝は、受検者の典型的な誤りを反映していた場合、受検者の能力水準に関する情報を一定程度もつことになる。受検者の能力に関する情報をもつ誤答選択枝は、能力の推定を行う上で有用なものになる (Bock, 1972). また、受検者の認知診断を行う上でも有用なツールとなりうる (Ozaki, 2015). 次節で詳細に述べるが、受検者の典型的な誤答を反映させた錯乱枝を提示することにより、受検者の能力が低いにもかかわらず正答できてしまうという状況を防ぐことができ、テスト項目の識別力を高めることに寄与する。

正答選択枝の作成にあたっては、客観的な事実や明確な解釈を利用することが一定程度可能であるため、誤答選択枝の作成に比べれば容易に作成できる。一方で、誤答選択枝の

作成にあたっては、正答選択枝の作成の際に利用した客観的な事実や解釈に至るまでの受検者の典型的な誤りを捉える必要があるため、正答選択枝の作成に比べても難しい。実際、Haladyna & Rodriguez (2013) においては、正答選択枝の作成に比べて、誤答選択枝の作成が難しいとの指摘もある。誤答選択枝の出来は、その項目の出来につながるものである(肥田野, 1972) ことを踏まえると、誤答選択枝をどのように作成するかということが、多枝選択式項目の品質を規定するものであると考えられる。

なお、誤答選択枝は一般に錯乱枝やまよわしとも呼ばれ(肥田野, 1972; Morimoto, 2008; 野口・大隅, 2014), いずれの用語もほぼ同義語として用いられることが多いが、本論文では誤答選択枝として位置づけたものの中に、受検者の典型的な誤りを反映していないものも存在するとの観点から、正答以外の選択枝全体を「誤答選択枝」と呼び、誤答選択枝のうち受検者の典型的な誤りを反映したものを「錯乱枝」と呼ぶ。錯乱枝の機能の詳細については、1.2.4 節で詳述する。

#### 1.2.4 錯乱枝の機能

以上に見たように、多枝選択式項目における選択枝、なかでも誤答選択枝の作成は多大な時間や労力がかかる一方で、作成された誤答選択枝の質の高さがテスト項目の品質に直結しうる。本節では、誤答選択枝がどのような観点からテスト項目の品質を高めるかについて、より詳細に述べることにする。なお、以降の記述においては、受検者の典型的な誤りを反映させた誤答選択枝のみに焦点を当てて述べるため、「錯乱枝」という用語を用いることにする。

多枝選択式項目における錯乱枝は、次の2つの機能をもつことが知られている。第1に、錯乱枝は項目の統計的性質を改善する機能をもっている。DiBattista & Kurzawa (2011) は、大学におけるさまざまな分野の講義で授業内に実施されているテストを材料に、受検者の典型的な誤答を反映させた誤答選択枝の機能を検討した。DiBattista & Kurzawa (2011) では、有効に機能していた錯乱枝の基準として、選択率が5%以上であること、錯乱枝の選択を1、非選択を0とした得点と合計得点との点双列相関係数が負の値を示すことの2点を挙げ、それぞれの項目で提示されている誤答選択枝のうち有効に機能していた錯乱枝の数と、その項目の正答率・点双列相関係数の関連について検討を行った。なお、錯乱枝に関する点双列相関係数は、合計得点が高くなる(能力が高くなる)ほどその錯乱枝を選択しないことが望ましいため、負の値をとることが期待されるものである(Attali & Fraenkel, 2000)。結果として、有効に機能する錯乱枝の数が少ない場合(0~1個)には、正答率が極端に高くなり、点双列相関係数が低い値であった一方で、有効に機能する錯乱枝の数が多い場合(2~4個)には、正答率が極端に高くなることはなく、点双列相関係数が高い値を示すことが明らかとなった。同様の結果は、看護学のテストを材料にした Tarrant et al.

(2009)においても得られている。DiBattista & Kurzawa (2011) や Tarrant et al. (2009) の知見は、項目の中に有効に機能する錯乱枝が多く存在することにより、適正な範囲の難易度の項目になると同時に、識別力の高い項目になることを示唆するものである。加えて、選択率が5%以上あり、かつ点双列相関係数が負の値を示す錯乱枝を作成することの重要性を示す研究知見であると言える。一方で、選択枝は3つ（錯乱枝は2つ）で十分であるとする知見も存在する (e.g., Rodriguez, 2005; Rodriguez et al., 2014)。

第2に、錯乱枝は認知診断の機能をもっている。Treagust (1988) は、理科のテストを題材にして、科学的現象に対する生徒の誤信念を明らかにするための認知診断テストを作成した。Treagust (1988) では、科学的現象そのものに関する項目と、その現象が生起する理由に関する項目をセットにした2層多枝選択式項目 (two-tiered multiple-choice format) を取り上げ、現象が生起する理由に関する項目において、生徒に対する認知的インタビューや自由記述式解答の結果に基づき受検者の誤信念を錯乱枝に反映させた。その結果、科学的な現象に対する知識についての問いでは90%程度の正答率を示していた一方で、現象が生起する理由についての問いの正答率は10%強であったことが明らかにされた (Treagust, 1988)。このようなことが明らかにできたのは、科学的現象の背後にある理由について受検者の誤信念を反映させた錯乱枝を用意したためであると推察される。受検者の典型的な誤答を含んだ錯乱枝を作成することにより、受検者の能力水準について、より詳細な推測を行うことが可能になるものと考えられる。なお、錯乱枝の認知診断的機能は、選抜の場面で用いられるようなハイスタークスのテストにおいても、テスト得点に対する適切な解釈を与えるという利点をもっていると考えられる。

なお、近年では認知診断モデル (cognitive diagnostic model) と呼ばれる一連のモデル群の開発と評価がさかんに行われている (e.g., de la Torre, 2009; Ozaki, 2015; 山口, 2016)。認知診断モデルとは、測定対象とする能力を構成する複数の下位スキルに分解し、下位スキルと項目への応答・選択枝への選択等の関係をQ行列として設定した上で、それぞれの下位スキルの習得状況を明らかにする統計モデルのことである。認知診断モデルの中には、錯乱枝に含まれる情報を認知診断に活用しようとするモデルも存在する (de la Torre, 2009; Ozaki, 2015)。大規模テストでの項目作成を念頭に置いている本論文では、錯乱枝の認知診断的機能を妥当性の向上という側面から捉えることになるが、認知診断を目的とするテストでは、錯乱枝の認知診断的機能を存分に生かし、受検者のスキルの習得の程度に関する推論を行うことが想定される。

このように、錯乱枝には項目の統計的性質を改善する機能と認知診断の機能があることがわかる。項目の統計的性質を改善する機能は定量的な側面からみた品質の向上に寄与し、認知診断の機能は定性的側面からみた品質の向上に寄与する。したがって、認知心理学の原則に基づいて、受検者の典型的な誤答を錯乱枝に反映させるという定性的な品質担保の手続きが、定性的な品質の向上はもちろん、定量的な品質の向上にも結びつくこと期待

される。本論文では、多枝選択式項目における錯乱枝を効果的に作成することが、テスト項目の定性的な品質を高めるだけでなく、定量的な品質の改善にも寄与するという点を踏まえ、受検者の典型的な誤答を反映させた錯乱枝について検討を行う。

上記の点を受け、1.3 節では、本論文が測定対象とする能力として、英語文章読解能力とその認知モデルをレビューし、受検者の典型的な誤答について検討を行う際の理論的基盤を概観する。

## 1.3 英語文章読解能力とその認知モデル

認知心理学の原則に沿って、受検者の典型的な誤答を反映した錯乱枝について検討を行うにあたり、測定対象とする能力を具体的に定める必要がある。本論文では、英語文章読解能力を測定対象として取り上げ、本論文での検討事項をより明確に述べることとする。

本節では、次の3つの点に触れながら、本論文がターゲットとする英語文章読解能力について整理する。1.3.1 節では、本論文において英語文章読解能力を取り上げる理論的意義・実践的意義について述べる。1.3.2 節では、本論文が焦点を当てる英語文章読解能力を2つの側面に分け、それぞれの特徴について述べた上で、両者の異同について説明する。1.3.3 節・1.3.4 節では、2つの英語文章読解能力のそれぞれに関する認知心理学の知見をレビューし、受検者の認知過程を概観する。本節の最後にあたる1.3.5 節では、前節までに行った英語文章読解能力に関するレビューを踏まえながら、英語文章読解能力を題材とした錯乱枝に関する先行研究の知見を概観する。

### 1.3.1 英語文章読解能力を取り上げる意義

本論文が英語文章読解能力を取り上げる理由は以下の2点である。具体的には、1) 英語を母語としない大学生にとって、英語で書かれた文章を読む能力が必須となりつつある時代にあること、2) 英語での文章読解は、認知心理学や第二言語習得の分野での知見の蓄積がなされており、受検者の典型的な誤答を捉える理論的な枠組みを提供しうること、の2点である。このうち、1点目は実践的意義に関連し、2点目は理論的意義に関連する。

英語文章読解能力を取り上げる第1の理由として、現代社会において、新聞記事・web ページ・学術論文・図書など、大学生の学習に必要な情報の多くが英語で書かれていることから、英語を母語としない大学生にとって、これらに記された情報を読解する能力が必須となりつつある点が挙げられる。自国以外にある情報に容易にアクセスできる現代において、母語以外の言語で書かれたテキストを読み、理解するという能力はますます重要度を増している (e.g., Khaki, 2014)。なかでも、英語で書かれたテキストを読むことは情報収集の中心となりつつあり、英語を母語としない大学生にとって必要なスキルである。第二言語としての英語学習を扱った多くの先行研究においても、英語文章読解能力は重要なス



キルとして位置づけられている (e.g., Lo et al., 2013). 本論文では、大学生において英語で書かれたドキュメントの読解能力が基礎的で重要なスキルとして位置づけられている点を考慮し、英語文章読解能力を取り上げることとした。なお、本論文では、外国語での文章読解の一例として英語を取り上げているにすぎず、英語文章読解能力の習得のみを肯定する立場には立っていない。

英語文章読解能力を取り上げる第2の理由は、認知心理学・第二言語習得の知見の蓄積が一定程度なされており、これらの研究知見が受検者の誤答を反映させた錯乱枝の作成に資する理論的な枠組みを提供しうる点である。次節以降で述べるように、Kintsch と van Dijk が文章読解の基礎的な認知モデルを提唱しており (Kintsch & van Dijk, 1978), このモデルが L2 学習者にも適用可能なモデルであることが示されて以降 (Johns & Mayes, 1990), 第二言語習得の領域で、多くの研究知見が蓄積されている (e.g., Buck et al., 1997; Keck, 2006; Kim, 2001). これらの領域における知見の蓄積は、受検者の認知プロセスを反映したテスト項目の作成に大きく貢献するものである。特に、認知モデルの中で仮定されているプロセスを考慮した設問を設定したり、各プロセスに特有の誤答の特徴を抽出したりすることで、受検者の典型的な誤答を反映させた錯乱枝の作成が可能になりうる。本論文において錯乱枝に関する検討を行うにあたり、認知モデルに基づいた受検者の典型的な誤答の把握が必要不可欠であることを考慮した結果、研究知見の蓄積が進んでいる英語文章読解能力を取り上げることとした。

### 1.3.2 英語文章読解能力の2つの側面

英語文章読解能力を測定するテスト項目について考える場合には、受検者が英語文章を読む際の認知プロセスを把握する必要がある。本節では、文章を理解するときの認知過程として、Kintsch (1998) の構築—統合モデル (construction-integration model) に基づき、読解力に関する整理を行った先行研究 (犬塚・椿本, 2014) および第二言語習得に関する先行研究を参考に、本論文が焦点を当てる英語文章読解能力の2つの側面について述べることにする。

構築—統合モデルでは、文章を読解したときに構築される意味の表象を、2つの軸から捉えている (犬塚, 2006; 犬塚・椿本, 2014; Kintsch & van Dijk, 1978; Kintsch, 1998). 1つ目は、部分—全体の軸であり、文章中の局所的な情報に関するボトムアップ処理を経て、文章全体の表象が構築されていく過程がイメージされている。この軸において、文章中に含まれる文中の局所的な情報処理の結果として構築される表象はマイクロ構造と呼ばれ、文章の全体的な意味処理の結果として構築される表象はマクロ構造と呼ばれる (Kintsch & van Dijk, 1978; Kintsch, 1998). 2つ目は、文章そのものの把握から知識と統合した理解に向かう軸であり、文章そのものから直接意味を引き出すような理解を経て、自らの知識と統合

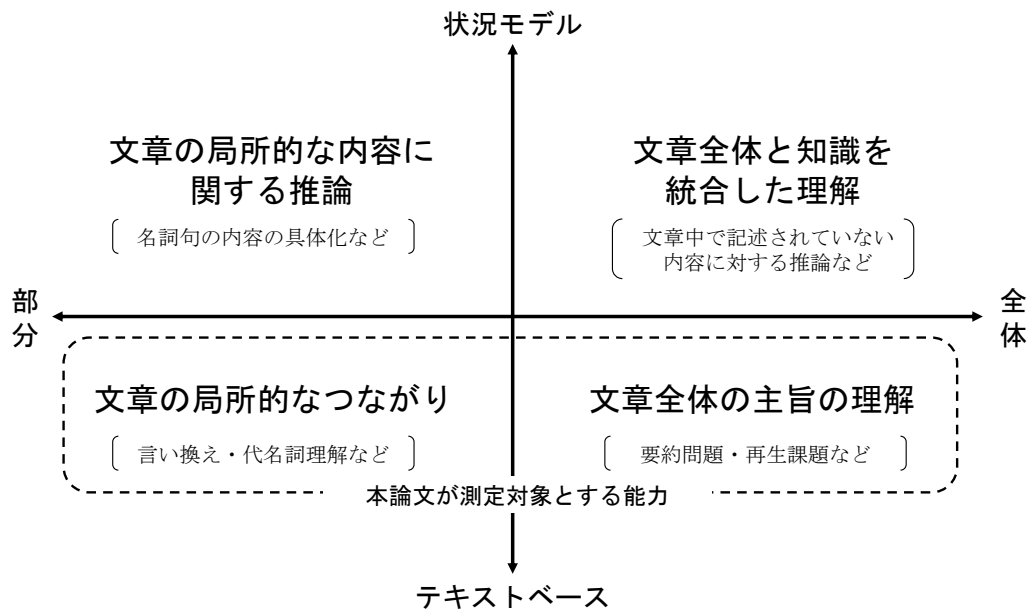


Figure 1.2. 構築—統合モデルにおける4種類の読解(犬塚(2006)をもとに著者が作成)と本論文が測定対象とする能力

された一貫性の高い表象を構築していく過程がイメージされている。この軸において、文章そのものに対する理解の結果として構築される表象はテキストベース、既有知識と結びついた表象は状況モデルと呼ばれる (Kintsch & van Dijk, 1978; Kintsch, 1998)。

これら2つの軸の組み合わせにより4種類の読解が想定され (Figure 1.2), 認知心理学ではそれぞれの読解を測定するための慣習的な方法が存在する (犬塚, 2006)。第1に、テキストベースにおけるマイクロ構造は、文章中の局所的なつながりを読解することに相当する。テキストベースのマイクロ構造を測定する際には、言い換えや代名詞の理解に関する問いを立てることが考えられる (犬塚, 2006)。第2に、テキストベースにおけるマクロ構造は、文章全体の主旨の理解に相当する。テキストベースのマクロ構造を測定する際には、情報の重要度について判断できた程度を把握する観点から、要約問題や文章の内容を再生する課題などが用いられる。第3に、状況モデルにおけるマイクロ構造は、文章中の局所的な内容に関する推論に相当する。具体的には、文章の内容に基づいてその後の結果を予測させる問題や、名詞句の内容を具体化させる推論を求める問題などが用いられる。第4に、状況モデルにおけるマクロ構造は、文章全体と既有知識を統合した理解に相当する。状況モデルのマクロ構造を測定する際には、文章中では記述されていない内容についての推論を求める問題を提示したり、文章の内容を再生する中で読み手が構築した状況モデルを特定する方法などがとられている。

以上に見るように、文章読解には2つの軸の組み合わせから構成される4つの側面が想

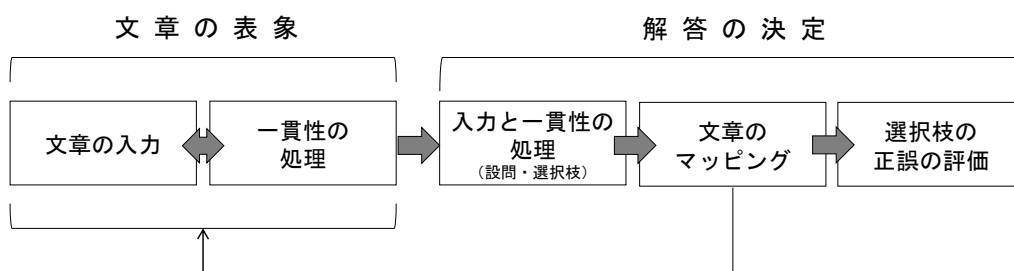


Figure 1.3. Embretson & Wetzel (1987) のモデル

定されている。本論文では、テストによる文章読解能力の測定という文脈を考慮し、文章そのものの内容理解の側面であるテキストベースの構築に焦点を当て、マイクロ構造とマクロ構造の両面から文章読解能力を捉えることとした。

### 1.3.3 英語文章中の局所的な情報処理に関する能力

前節で見たように、テキストベースのマイクロ構造の構築とは、文章中の句や節、文などの局所的な部分の理解に相当する過程である。特に、受検者が文章読解問題に解答する場合には、問いの内容に関連する文章中の局所的な情報を特定する過程が想定される。Embretson & Wetzel (1987) は、文章読解問題を解く際の情報処理に関するモデルを提唱している (Figure 1.3)。このモデルでは、文章読解問題を解く際の情報処理として、文章の表象を構築する過程と、解答を決定する過程の2つのプロセスが仮定されている。

文章の表象を構築する過程では、さらに下位の2つのプロセスが想定されており、文章中の文の入力 (text encoding) と一貫性の処理 (coherence process) から構成されている。まず受検者は、提示された文章中の文を1つ1つ処理し、マイクロな命題を構築する。複数のマイクロな命題が構築された後、それらが一貫した命題群となるように、命題間の結びつきに関する処理を行ったり、命題そのものの再構築などを行う。一貫した命題群を構築するにあたり、文章中の関連する文を読み返したり、命題間の結びつきを推測したりする認知活動を行っている。

文章の表象が構築されたのち、受検者は文章読解問題の設問の部分を読み、解答の決定を行う。解答を決定する過程においては、入力と一貫性の処理、文章のマッピング、選択枝の正誤の評価という3つの下位プロセスが想定されている。1つ目は、設問と選択枝を入力し、一貫性の処理を行うプロセスであり、提示された文章に対する処理と同様の処理が、設問幹・選択枝に対しても行われる。2つ目は文章のマッピングの過程であり、提示された選択枝が正しいことを確認したり、誤っている箇所を特定したりするために、それぞれの選択枝に関連する文章中の命題を見つけ出す過程である。3つ目は選択枝の正誤の評価の過程であり、文章のマッピングの過程で参照した命題と選択枝を照らし合わせながら誤

りを特定していく過程 (falsification), 残った選択枝の正しさを確認する過程 (confirmation) である。受検者は、これらの過程を経て正答だと思う選択枝を決定することになる。

上記のモデルのうち、解答の決定の段階における3つの下位プロセスは、文章読解問題を解く場合に特有のものであると考えられる。解答の決定の段階では、設問や選択枝を読み、それぞれの選択枝の内容に関連する文章中の命題（記述）を特定した上で、選択枝の内容に関連する命題と選択枝そのものを比較し、選択枝中に含まれる誤りを見つけたり、選択枝の正しさを評価したりする認知処理が仮定されている。実際 Embretson & Wetzel (1987) では、解答の決定の過程に関連する指標 (e.g., 選択枝に関連する内容が文章中に含まれる割合、選択枝中の誤りを特定する過程など) が項目困難度を説明することが示されている (Embretson & Wetzel, 1987; Embretson, 2001)。また、L2 学習者を対象として TOEIC で使用された項目を材料に検討を行った Buck et al. (1997) においても、文章中の記述と選択枝を見比べて誤りを見つける段階に関連する能力と合計得点の間の正相関が報告されている。

以上を踏まえると、文章中の局所的な理解を求める項目では、設問や選択枝の内容に関連する記述の特定（文章中のマッピング）や、その記述と選択枝の比較（選択枝の正誤の評価）の過程が、受検者の正誤を決定づけているものと考えられる。

#### 1.3.4 英語文章の要点を把握する能力

1.3.2 節で見た通り、テキストベースのマクロ構造は、文章全体の理解に相当するものであり (Kintsch & van Dijk, 1978), 認知心理学では要約課題などがよく用いられる (犬塚, 2006)。文章全体の命題表象を構築するときには、ミクロ構造として構築された命題群のうち重要でないものを取り除き、重要な情報に注意を向けながら読解することが必要である (Brown & Day, 1983)。

Kintsch & van Dijk (1978) では、以下の3つのマクロルールに沿って、構築された命題を取捨選択・統合する (Kintsch & van Dijk, 1978)。第1に削除 (deletion) という過程で、文章のトピックに関連が薄いと考えられる命題を取り除き、第2に一般化 (generalization) という過程で、命題群をより上位概念・上位集合の表現に置き換えた上で、第3に構築・統合 (construction/integration) という過程で、ミクロ構造の命題群を中心に据えながら、より全体的な命題を構築する。文章の読み手は、上記3つのマクロルールを適用しながら文章中の命題を取捨選択して読解することが知られている。

Brown & Day (1983) は Kintsch & van Dijk (1978) のマクロルールを要約文作成の過程として拡張・精緻化し、次の6つの認知過程を仮定している。具体的には、a) 重要でない情報、および b) 冗長な情報を削除し、c) 並列されたことばや d) 動作を上位概念の語に置き換えた上で、e) 文章の筆者によるトピックの要約（トピック・センテンス）がある場合

はトピック・センテンスを見つけ、f) トピック・センテンスがない場合は自らトピック・センテンスを作るとしている (Brown & Day, 1983). このモデルの a と b は Kintsch & van Dijk (1978) のモデルにおける削除に対応し、c と d は一般化、e と f は構築・統合に対応する。

以上に見るように、文章の要約の際、削除・一般化・統合というマクロルールを適用することが示されている。Kintsch & van Dijk (1978) や Brown & Day (1983) が提唱した認知過程は、母語で書かれた文章を読む際のもので提唱されたものであるが、L2 学習者が英語で書かれた文章を読んだり、要約したりする際の過程としても適用可能であるということが示されている (Johns & Mayes, 1990; Keck, 2006; Kim, 2001). Kintsch & van Dijk (1978) および Brown & Day (1983) の枠組みが L2 学習者の読解にも適用可能であることは、L2 学習者における英語文章読解項目の作成の参考資料として活用できることを意味している。なお、L2 学習者におけるマクロ構造の構築の詳細は、4.1 節で述べることとする。

### 1.3.5 英語文章読解項目における錯乱枝についての先行研究とその課題

認知心理学に関する先行研究では、さまざまな観点から英語文章を読む際の特徴が明らかにされている。一方で、英語文章読解項目における錯乱枝についての先行研究では、主にテキストベースのマイクロ構造の側面の英語文章読解能力に焦点が当てられてきた。本節では、英語文章読解項目における錯乱枝についてのこれまでの研究知見を概観しながら、先行研究の課題およびその解決方策について述べることとする。

英国のテスト企業のひとつである Pearson Assessment Inc. は、英語文章読解項目における錯乱枝の理論的分類 (Distractor Rationale Taxonomy) を提唱している (King et al., 2004; Lin et al., 2010). 錯乱枝の理論的分類では、受検者の文章読解上の誤りのレベルを Level 1 から Level 4 の 4 段階に分けており、それぞれのレベルの錯乱枝が受検者のどのような誤りを反映したものであるかを示している。4 段階のうち、文章中の内容とは関連しないことがらに焦点を当てた錯乱枝が Level 1 に、文章中に記述されている内容どうしの関係を取り違えた錯乱枝が Level 2 に、文章の内容に対する正しい分析や解釈を含んでいるが、結論や推論でいくらか正答選択枝に劣る錯乱枝が Level 3 に位置づけられ、正答選択枝を Level 4 としている。また、Lin et al. (2010) は順序多枝選択式項目 (ordered multiple-choice items) を用いて錯乱枝の理論的分類を各選択枝に反映させ、従来の多枝選択式項目よりやや高い信頼性が示されたこと、外的基準との相関が一定程度高いこと、順序多枝選択式項目のほうがより受検者の理解状況に関する情報をもっていることなどを報告した上で、錯乱枝の理論的分類の妥当性を実証的に確認している。Lin et al. (2010) の結果は、錯乱枝の理論的分類を踏まえた多枝選択式項目の作成の有用性を示唆するものである。

さらに、第二言語習得の分野では、Ushiro et al. (2007) が TOEFL PBT の多枝選択式項目

を材料に、錯乱枝に関する検討を行っている。その結果、英語文章中のキーワードを含む錯乱枝や、一般性が高く多くの情報を含む錯乱枝などが多くの受検者に選ばれていたことを明らかにしている (Ushiro et al., 2007)。このように、錯乱枝の質的な違いに着目した研究知見は、錯乱枝の作成に有用であると考えられる。

これらの先行研究は、錯乱枝の作成に大きく寄与する一方で、以下の3つの課題があるものと考えられる。第1に、錯乱枝に反映されている受検者の典型的な誤答に関する知見として具体性に欠ける点である。例えば、King et al. (2004) や Lin et al. (2010) が依拠する錯乱枝の理論的分類では、文章中に記述されている内容どうしの関係を取り違えた錯乱枝が Level 2 に位置づけられているが、具体的にどのようなものの取り違えを反映させる必要があるかという点について、十分な示唆が得られていないものと考えられる。また、Ushiro et al. (2007) では、英語文章中のキーワードを含む錯乱枝や、一般性の高い情報を含む錯乱枝の効果が示唆されているが、英語文章中のキーワードの難易度に依存するかどうかといった点や、一般性の高い情報を含めた選択枝に受検者の典型的な誤答をどのような形で反映させるかという点など、項目作成の段階で即効性の高い研究知見であると見なすことは難しいものと考えられる。先行研究の状況に照らして、より具体的な受検者の誤答に関する知見を提供することにより、錯乱枝の効果的な作成に資することが期待される。

第2に、受検者の能力水準ごとにみた検討がなされていない点である。英語文章読解能力を題材としたものではないが、英文法の知識を測る空所補充形式の項目を用いて錯乱枝に関する検討を行った Morimoto (2008) は、受検者の能力水準ごとに錯乱枝の選択状況に関する検討を行い、能力水準の低い受検者には意味的に関連のない語を提示した錯乱枝が選択されやすく、能力水準の高い受検者には統語的な誤りを反映した語を提示した錯乱枝が選択されやすいことを明らかにしている。このように、錯乱枝の選択状況を能力水準ごとに検討することは、錯乱枝の特性を知るために有効な方法のひとつである。英語文章読解項目における錯乱枝の特性を能力水準ごとに把握することにより、項目作成に対する示唆を得ることができる可能性がある。

第3に、先行研究では既存の標準テストにおける錯乱枝を事後的に分類したものが多く、受検者の典型的な誤答を踏まえた錯乱枝の知見とは見なしにくい点である。たとえば Ushiro et al. (2007) では TOEFL PBT のテスト項目を利用し、事後的に錯乱枝の分類を行っている。このような方法は、すでに作成されたテストにおける錯乱枝の特徴を把握する場合には適しているが、認知心理学の原則に基づく項目作成の立場に立てば、項目作成の段階で受検者の典型的な誤答を錯乱枝に反映させる手続きが必要であると考えられる。なお、テスト項目の特性に関する検討にあたって、研究目的に沿ったテスト項目の作成を行っているものに Ascalon et al. (2007) があるが、自動車の運転に関する知識を問うテストを題材としたものであり、英語文章読解能力を測定するテストを題材としたものは見当たらない。

そこで本論文では、英語文章読解能力を測定する多枝選択式項目において、どのような誤答を反映させた錯乱枝が、どの能力水準で選ばれやすいのかについて、研究目的に沿ったテスト項目を作成することにより検討を行う。このことにより、上に挙げた先行研究の課題3点を解決しながら、英語文章を読解する受検者の典型的な誤答を反映させた錯乱枝の作成手法に関する具体的知見を提供することが可能になるものと考えられる。

## 1.4 本論文の目的と構成

### 1.4.1 本論文の目的

1.1節から1.3節までの整理を踏まえ、本論文では、多枝選択式の英語文章読解項目における錯乱枝の作成に役立てることを念頭に入れ、受検者の典型的な誤答を選択枝に反映させた場合の錯乱枝の選択率について検討を行うことを目的とする。なお、錯乱枝の機能を示す統計的指標として選択率を取り上げた背景には、各研究の受検者数の違いに左右されずに、一貫した解釈を行うことができる点が挙げられる。

本論文では、多様な英語文章読解能力のうち、英語文章の局所的な読解を行う側面と、大局的な読解を行う側面の2種類に焦点を当て、それぞれの認知過程を踏まえながら、受検者の典型的な誤答の特徴を明らかにする。その上で、典型的な誤答の特徴を反映させた錯乱枝の選択率について、能力群ごとにみた検討を行い、それぞれの錯乱枝の有効性に関する示唆を得る。

### 1.4.2 本論文の構成と各章の概要

本節では、本論文における各章の概要を説明しながら、本論文の構成について述べる。Figure 1.4に、本論文の構成を表す概念図を示した。なお、Figure 1.4の矢印は本論文における論の流れを表しており、矢印の始点に位置づけられる章の議論や研究知見に基づいて、矢印の終点に位置づけられる章の議論や研究が展開されることを示している。

本論文は、全5章から構成される。本章(第1章)では、本論文における序論として、先行研究の整理を行い、本論文の着眼点・目的などについて述べた。本章で整理された研究知見を踏まえ、後続する3つの章(第2章～第4章)では、多枝選択式の英語文章読解項目における錯乱枝の選択率に関する実証的研究を行う。本論文の構成を表したFigure 1.4では、本章から第2章・第3章に続く流れと、本章から第4章につながる流れの2つに分岐している。第2章・第3章は、本論文が焦点を当てる英語文章読解能力のうち、ミクロ構造的な側面に焦点を当て、第4章は、マクロ構造的な側面に焦点を当てる。第5章では、第2章・第3章において得られた知見と、第4章において得られた知見を整理し、多枝選択式の英語文章読解項目を作成する際の示唆や今後の展望について述べる。以下では、各

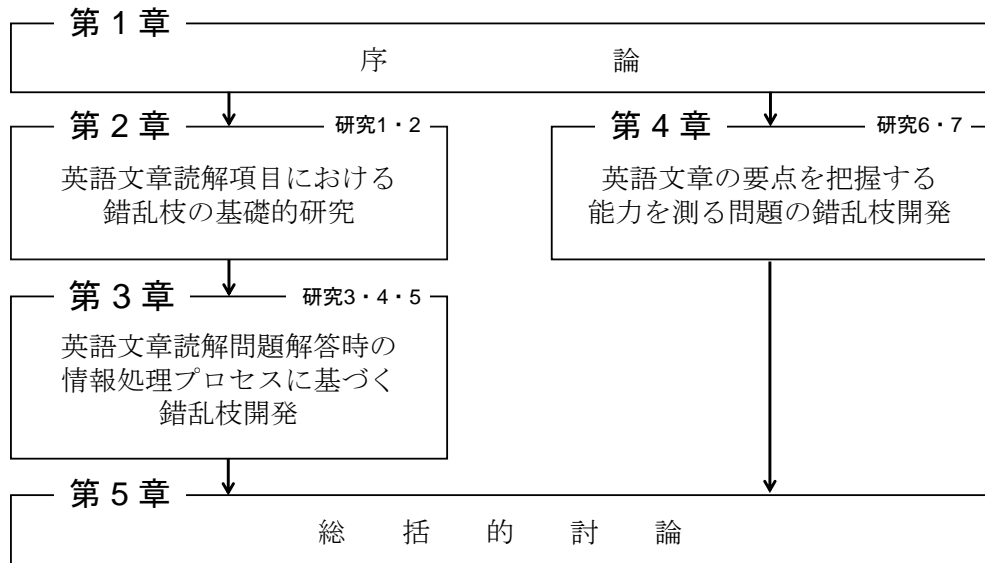


Figure 1.4. 本論文の構成

章の概要を紹介する。

ミクロ構造的な側面の読解能力について扱った章のうち、第2章では、Embretson & Wetzel (1987) が提唱したモデルにおける「解答の決定」のプロセスの中で、選択枝の正誤評価の過程に焦点を当て、どのような特徴をもつ錯乱枝が存在し、どのような能力水準にある受検者に選択されやすいかということを示す。2.2節では研究1として、過去に実施された大学入試問題を材料とし、受検者の典型的な誤答を抽出する。2.3節では研究2として、研究1の結果を踏まえ、抽出された受検者の典型的な誤答を選択枝に反映させ、受検者の能力群ごとに選択率に関する検討を行う。

また、第3章では、選択枝の正誤の評価の過程に加え、文章のマッピングの過程も考慮に入れ、錯乱枝の選択率に関する検討を行う。具体的には、設問や選択枝を読んだ後に、各選択枝の内容に関連する英語文章中の記述 (i.e., キーセンテンス; Carr, 2006) を特定する認知処理や、英語文章中の記述を選択枝と見比べて両者の違いを検討する認知処理が、設問設定によって規定される可能性があることを念頭に入れ、設問の作り分けによってこれらの認知処理に関する実験操作を行った。3.2節では研究3として、設問においてキーセンテンスの特定の容易さを操作するための予備調査として、キーセンテンスの位置関係に関する調査を行う。3.3節では研究4として、研究3の結果を踏まえ、キーセンテンスの位置関係を設問において操作し、研究2で独特の効果が示された錯乱枝3種類について、受検者の能力群ごとに選択率に関する検討を行う。3.4節では研究5として、キーセンテンスと選択枝の比較の容易さを操作するため、キーセンテンス中の語を含んだ条件と、



キーセンテンス中の語の類義語を用いた条件の錯乱枝を作成し、受検者の能力群ごとに選択率に関する検討を行う。

第4章では、英語文章読解能力のうちのマクロ構造的な側面に着目し、英語文章の要点を把握する能力を測る多枝選択式項目に焦点を当てて、錯乱枝の選択率に関する検討を行う。英語文章の要約に関する認知モデル(e.g., Brown & Day, 1983; Kintsch & van Dijk, 1978)に基づき、削除・一般化・統合のプロセスにおける受検者の誤答パターンを明らかにした上で、受検者の誤答パターンを選択枝に反映させ、受検者の能力群ごとに選択率に関する検討を行う。4.2節では研究6として、英語文章の要点把握における受検者の誤答パターンを見出すため、記述式の要約問題に対する解答から受検者の誤答パターンを抽出する。4.3節では研究7として、研究6の結果を踏まえ、英語文章の要点把握における受検者の典型的な誤答を反映させた選択枝のそれぞれについて、能力群ごとに選択率の検討を行う。

第5章では、5.1節で本論文から得られた研究知見を整理し、5.2節で本論文における研究の意義や今後の課題を述べる。5.3節では項目作成に関する研究の今後の展望などについて述べ、5.4節で本論文を締めくくる。



## 第2章

# 英語文章読解項目における錯乱枝の基礎的研究

### 2.1 研究の背景

本章では、英語文章読解問題における錯乱枝の機能を把握するための基礎的研究として、錯乱枝に反映されている誤答の特徴を洗い出し、その誤答の特徴を反映させた錯乱枝を実験的に作成して、選択率に関する検討を行う<sup>1</sup>。

英語テストにおける錯乱枝の機能について検討を行った先行研究の多くは、語彙能力を測る空所補充問題を材料としたものである (Goodrich, 1977; Morimoto, 2008)。英文法問題では単語そのものが錯乱枝となりうるが、Goodrich (1977) においては、文脈に関連した語や対義語、類義語を錯乱枝とした場合に、選択率が高かったことが報告されている。また、Morimoto (2008) は能力水準別に錯乱枝の選択率の検討を行い、能力低群では無関連語を錯乱枝としたものの選択率が高かった一方で、能力高群では意味的に正しいが統語的な誤りを含んだ語を錯乱枝としたものの選択率が高かったことを明らかにしている。

これらの先行研究は英文法問題を材料としたものであるが、英語文章読解問題を材料としたものはほとんど見られない (King et al., 2004; Lin et al., 2010)。King et al. (2004); Lin et al. (2010) は、英語文章読解問題における錯乱枝の理論的分類 (Distractor Rationale Taxonomy) を提唱し、英語文章読解問題の選択枝をレベル別に4分類した。具体的には、Level 1 として「文章に関連しない内容を含む錯乱枝」、Level 2 として「文章中に記述されている事実間の関係を取り違えている錯乱枝」、Level 3 として「分析や解釈を必要とし、正答選択枝に比べるといくらか弱い錯乱枝」があり、最後に Level 4 として「正答選択枝」を設定している。

多枝選択式問題の作成の際、錯乱枝の理論的分類に基づいて錯乱枝を作成することは非常に有用であり、テスト項目の品質を定量的・定性的に高めることに資するものと考えら

---

<sup>1</sup> 本章は、寺尾他 (2015) を加筆・再構成したものである。

れる。その一方で、錯乱枝の理論的分類におけるそれぞれのレベルの錯乱枝にどのようなものが存在するのか、実際にそのような錯乱枝は受検者の能力を識別するのに有用であるか等、実証的な研究知見が現段階では蓄積されていないと言える。

錯乱枝の有用性を示すためには、1) 錯乱枝がどのような特徴をもっているか、2) 能力水準ごとにみた錯乱枝の選択率がどのようにになっているかという2点について検討を行う必要がある。そこで本章では、これら2つの課題点を解決するため、第1の研究課題を研究1(2.2節)で、第2の研究課題を研究2(2.3節)で検討する。研究1では、過去の入学試験問題における英語文章読解問題を材料とし、誤答選択枝に反映されている誤答の特徴を洗い出す基礎的研究を行う。研究2では、研究1で明らかにされた誤答の特徴を錯乱枝に反映させた多枝選択式問題を作成し、大学生の受検者に解答を求めた上で、錯乱枝の選択率を能力群ごとに検討する。研究1により上記の課題点1(錯乱枝がどのような特徴をもっているか)に関する知見を得た上で、研究2を行い、上記の課題点2(能力水準ごとにみた錯乱枝の選択率の様相)に関する知見を得る。これらの知見は、問題作成者が多枝選択式問題を作成する際のガイドラインとなり得るものであり、テスト項目の品質を向上させる上で意義のある知見であると考えられる。

## 2.2 研究1：多枝選択式問題における錯乱枝の特徴の把握

研究1では、錯乱枝の有用性を実証的に示すための第1の課題として挙げた、英語文章読解問題の錯乱枝に反映された誤答の特徴の把握を目的とする。

### 2.2.1 方 法

#### 2.2.1.1 研究協力者

大学生・大学院生16名(男性3名、女性13名)を研究協力者とした。本研究の研究協力者は、難関国立大学に通う者であるため、英語文章読解能力が高いと考えられ、英語文章読解問題における誤答選択枝の誤答の理由を正しく指摘するという本研究の目的を達成することが可能であると判断される。

#### 2.2.1.2 材 料

インターネット上で公開されている大学入学試験問題の英語テストのうち、多枝選択式の文章読解問題を11題選定した。英語文章読解問題はすべて、問いと選択枝の両方が英語で書かれているもの(英問英答)であり、計45個の設問を抽出した。

大学の入学試験問題は、センター試験・国公立大学の個別学力試験・私立大学の個別学力試験に大別されるが、本研究では私立大学の個別学力試験を用いることとした。この理由として、私立大学の個別学力試験では多枝選択式問題が多く用いられており、英語文章

Table 2.1  
研究 1 で使用した調査冊子・文章・設問数一覧

冊子	文章 No.	テーマ	語数	リーダビリティ	設問数
冊子 1	1	二酸化炭素価格	743	10.87	6
冊子 2	2	アメリカとヨーロッパの交通事情	614	9.05	5
冊子 3	3	ジョナサン・ハイトの道徳理論	569	10.11	3
冊子 4	4	権力のシフト	1156	8.84	7
冊子 5	5	陪審員制度	884	13.54	7
冊子 6	6	ディープエコロジー	196	7.72	2
	7	タトゥー	185	12.58	1
冊子 7	8	想像上の観衆	284	14.64	3
冊子 8	9	研究	455	10.39	5
冊子 9	10	シラノ・ド・ベルジュラックの一生	259	10.96	2
冊子 10	11	乳児期における他者とのかかわり	492	11.12	4

読解にあたっての多様な誤答が選択枝に反映されていることが挙げられる。実際、私立大学の入学試験問題は、出題大学・学部によって、選択枝に反映されている誤答のレベルも異なることが想定される。一方で、センター試験の問題は比較的易しく、選択枝に反映されている誤答の特徴が限定的になる可能性があること、国公立大学の個別学力試験には多枝選択式問題よりも記述式問題が多いことから、これらの入学試験を材料として取り上げることとはしなかった。本研究では、可能な限り多様な誤答選択枝を提示し、さまざまな誤答の特徴を収集することを目的としているため、多様な誤答が反映されていると判断される私立大学の入学試験問題を使用した。

英語文章読解問題の選定にあたっては、以下の2つの基準を設けた。第1に、入学試験においてさまざまな難易度の文章が提示されている実状に照らして、極端に易しい文章を可能な限り避けながら、文章の難易度にバリエーションをもたせた。本研究において英語文章の難易度を揃えた場合、特定の難易度の英語文章をもつ読解問題のみを対象として誤答の特徴に関する知見を提供することとなる。このような知見は、入学試験において英語文章の難易度にバリエーションがあるという実状とは異なる状況での検討となることに加え、抽出できる誤答の特徴の多様性を損なうことになる可能性がある。本研究では、さまざまな難易度の英語文章を用いて誤答の特徴を把握することにより、多様な誤答の特徴を抽出できるようにした。

本研究では、使用候補としたそれぞれの英語文章について、文章の難易度に関する指標として知られているリーダビリティをあらかじめ算出し、英語文章の難易度を事前に確認した上で、極端に難易度の高い文章を使用しないこととした。文章のリーダビリティにはさまざまな指標があるが、本研究では、大学入学試験問題を材料に英語文章のリーダビリティについて検討を行った Brown & Yamashita (1995) や Kikuchi (2006) を参考に、Flesch-

Kincaid の公式を用いて文章のリーダビリティを算出した。Flesch-Kincaid の公式は、総語数を総文数で割った「文の長さ」と、総音節数を総語数で割った「語の長さ」を基に算出されるリーダビリティの指標であり、その文章がアメリカの学年にして何年生向けであるかを示している。本研究では、8年生（わが国における中学2年生）レベルの文章から14年生（大学2年生）レベルの文章までを網羅するよう、英語文章を選定した。Table 2.1 に、研究1で使用した文章のテーマと語数・リーダビリティとともに、各文章に付随していた設問の数を示した。

第2に、文章の内容理解に関する問題のみを抽出し、選択枝が句や節、文で提示されている問題を選定した。英語文章読解問題には、文章中の語の意味を問うものや、文章中の文を並べ替えるものなどもあるが、本研究では英語文章読解問題中の語彙問題や文法問題は取り扱わず、英語文章の内容理解問題のみを検討対象とした。

### 2.2.1.3 調査冊子の構成

上記の基準で選定された英語文章読解問題11題（英語文章・設問の幹・選択枝を11セット）を、計10種類の調査冊子に収録した（Table 2.1）。1種類の調査冊子につき、原則1題の英語文章読解問題を収録したが、語数の少ない英語文章をもつ読解問題に関しては、2題まとめて1種類の調査冊子に収録した。

1種類の調査冊子では、英語文章・設問・誤答の理由を指摘するためのコメント欄から構成された。1つの英語文章につき、2～7問の設問を提示した。研究協力者には、英語文章・設問および正答選択枝1つと誤答選択枝3つの計4つの選択枝を提示し、正しいと思う選択枝の番号を選ぶよう求めた。次に、研究協力者が選ばなかった選択枝のそれぞれについて、誤答だと判断した根拠となる記述が英語文章中にあれば、その部分に下線を引くよう求めた。さらに、選択枝の下に提示したコメント欄に、研究協力者が選ばなかった選択枝のそれぞれについて、誤答だと思う理由を記述するよう求めた。なお、正答だと思った選択枝に対しては、コメントを記入しないよう教示した。Figure 2.1 に、設問・選択枝・コメント欄の例を示した。

### 2.2.1.4 手続き

調査は、2013年1月～3月の間に行われ、実験室において個別に実施された。研究協力者には、この調査における解答・回答が研究協力者自身を評価するものではないこと、調査の途中で解答・回答したくないと思った場合には、調査協力を中断できることなどを伝えた。本研究の目的について十分説明を行った上で、研究協力者の同意が得られた回答データのみを分析に含めた。研究協力者には、調査の開始前に回答例を用いて回答の仕方について説明した。

まず、調査冊子への回答にあたり、研究協力者には、英語文章を読んで、設問ごとに正

問1 Why does the author refer to Australian Rules football?

- ① Because climate change is as difficult to understand as Australian sports.
- ② Because those who deny climate change don't really understand science.
- ③ Because climate change has been caused primarily by human activity.
- ④ Because there are only a few people who truly understand climate change.

①
②
③
④

Figure 2.1. 研究1 で使用した調査冊子の例（設問・選択枝・コメント欄）

答だと思える選択枝の番号を塗りつぶすよう教示した。次に、正答でないと思った各選択枝について、選択枝中で誤りだと思える箇所に下線を引き、誤答の理由をコメント欄に記述するよう求めた。回答時間は設けず、研究協力者がすべての設問への解答・回答を終了した時点で調査を終了した。

調査実施にあたり、まず10種類の調査冊子のそれぞれに対して各1名の研究協力者から解答・回答を求め、調査冊子に対する解答・回答の状況を確認した上で、誤答の情報の収集が十分でないと判断された調査冊子については、もう1名の研究協力者に解答・回答を求めた。追加で解答・回答を求めた調査冊子は6種類あったため、追加で6名の研究協力者に協力を依頼した。

## 2.2.2 結果と考察

### 2.2.2.1 誤答の理由に関する集計結果

英語文章11題・設問45問に対して、誤答選択枝に反映された誤答を指摘したコメントが計206件得られた。誤答に対して得られたコメントを概観すると、研究協力者2名を割り当てた調査冊子において、同一の選択枝であっても異なる種類の誤答を指摘していた

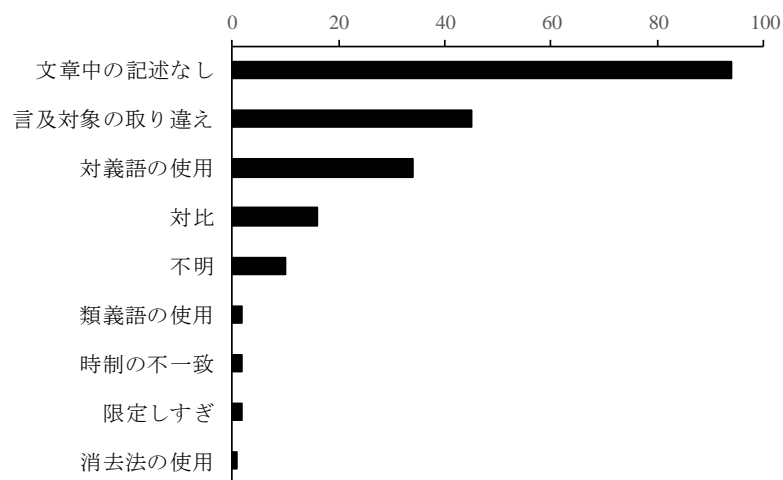


Figure 2.2. 誤答の理由に関するコメントの集計結果

ケースが多く見られた。本研究では、誤答の多様性を捉えることが目的であること、どちらの研究協力者のコメントも誤答の理由として妥当な指摘であったことから、提示した誤答選択枝が複数の種類の誤りを含んでいると判断し、両方をカウントした。また、正答選択枝に対して誤答の理由を指摘していたコメントは、カウントに含めなかった。上記のようなカウント方法に基づいて、誤答に関する指摘をカテゴリごとに集計した。コメントの集計結果を Figure 2.2 に示した。以下では、Figure 2.2 に基づいて、具体的なコメントを取り上げながら、度数の多かった順に詳細を述べていくこととする。

誤答に対するコメントとして最も多く得られたのは、英語文章中では記述されていないことがらに誤答選択枝中で言及されているとの指摘であり、類似した指摘が 94 件得られた。そこで、これらのコメントに対するカテゴリを「文章中の記述なし」と命名した。

2 番目に多かったのは、設問で問われていることに対する答えとして、選択枝中で言及されている内容や因果関係、比較対象を取り違えて記述しているような誤答を指摘するコメントであり、計 45 件のコメントが得られた。このようなコメントに対するカテゴリを「言及対象の取り違え」と命名した。このカテゴリは、当初「内容の取り違え」「因果関係の取り違え」「比較対象の取り違え」という独立した 3 カテゴリであったが、英語文章中の何らかを取り違えているという点で共通点があると解釈し、1つのカテゴリにまとめることとした。以下では、それぞれの下位カテゴリ（内容の取り違え・因果関係の取り違え・比較対象の取り違え）のそれぞれについて説明する。

言及対象の取り違えに関する 1 つ目の下位カテゴリ「内容の取り違え」は、英語文章中



の記述の名詞部分を取り違えたような誤答として解釈できる。例えば、文章中に「ニューヨークの交通は自動車中心である」と述べられていた場合に、誤答選択枝として「ニューヨークの交通は歩行者中心である」が提示されていた。このような状況において、研究協力者から「歩行者中心ではなく自動車中心だから」のようなコメントが得られた。

言及対象の取り違えに関する2つ目の下位カテゴリ「因果関係の取り違え」は、順接や逆接、原因と結果を取り違えたような誤答として位置づけられる。例えば、文章中に「情報技術が発展するほど、情報技術が普及していく」という記述があった場合に、誤答選択枝では「情報技術の普及が、情報技術の発展をもたらす」と述べられていた。このような状況において、研究協力者から「因果関係が逆である」のようなコメントが得られた。文章中では、情報技術の発展を原因とし、情報技術の普及を結果として述べられていたにもかかわらず、誤答選択枝中では原因と結果を取り違えて記述していることになる。

言及対象の取り違えに関する3つ目の下位カテゴリ「比較対象の取り違え」は、主に二つ以上のもを比較するような内容の英語文章で見られた誤答のカテゴリであり、AとBを比較すべきところを、AとA'の比較として記述しているような誤答である。例えば、アジアとヨーロッパを比較しながら論旨が展開されている文章を提示した場合に、付随する設問の中には、以前のアジアと現在のアジアの特徴を捉えるような設問が含まれていた。この設問において、ヨーロッパに関する記述を含む誤答選択枝に対して、「比較対象は以前のアジアであってヨーロッパではない」のようなコメントが得られた。

3番目に多かったのは、誤答選択枝中で英語文章中の記述の対義語を用いた誤答であり、計34件のコメントが得られた。例えば、英語文章中で“short”と記述されているにもかかわらず、誤答選択枝中で“long”と記述されていた場合、研究協力者は“long”に下線を引き「文章中にはshortとある」とコメントしていた。このような誤答に対するカテゴリとして「対義語の使用」と命名した。

4番目に多かったのは、文章全体の対比構造(e.g., 政府 vs. 国際的組織)に焦点を当てた誤答であり、誤答選択枝中で一方の特徴を他方の特徴として記述しているような誤答であった。これに類するコメントは、計16件得られた。例えば、国際政治に関する英語文章中で、政府を超越するような大きな国際的組織の必要性が述べられており、設問では国際的組織の特徴に関する問いが立てられていたにもかかわらず、国際的組織の特徴に関する記述文が政府の特徴にすり替わっているような誤答選択枝が提示された場合に、研究協力者から「政府を超越するような組織が主題である」のようなコメントが得られた。

5番目の「不明」は、「なんとなく誤答」のようなコメントであり、誤答の理由に対する指摘として明確でなかったものをまとめたものである。6番目の「類義語の使用」は、文章中の語と誤答選択枝中の語が類似した意味であることを指摘したものであった。また、7番目の「時制の不一致」は、文章中では未来形で書かれていたものが、誤答選択枝中では過去形で書かれていることを受け「これからの話」などと指摘していたものであった。

Table 2.2  
研究 2 で扱う錯乱枝の要因・水準とその定義

要因	水準	定義
文章中の記述の有無	文章中の記述あり	当該錯乱枝の記述を誤りとする内容が文章中に記述されている
	文章中の記述なし	当該錯乱枝の記述を誤りとする内容が文章中に記述されていない
対概念の使用	否定語の使用	当該錯乱枝に, <b>not・no・few・little・un-</b> などの否定語・否定辞を含んでいる
	対義語の使用	当該錯乱枝に, 文章中に使用されている語の対義語を含んでいる
	対比関係の使用	当該錯乱枝に, 文章全体の対比関係を利用した表現を含んでいる
言及対象の取り違え	内容の取り違え	文章中の内容を取り違えて表現している
	因果関係の取り違え	文章中の因果関係を取り違えて表現している
	比較対象の取り違え	文章中の比較対象を取り違えて表現している

8 番目の「限定しすぎ」に関しては, 文章中の内容の一部だけに限定して記述されていた誤答選択枝に対して「〇〇に限定されないから」のような指摘であった. 9 番目の「消去法の使用」に関しては, 「消去法で」のようなコメントであり, 5 番目の「不明」と同様, 誤答選択枝に反映されている誤答を明確に指摘したものではなかった.

### 2.2.2.2 研究 2 で取り上げる錯乱枝

研究 2 では, 効果的と思われる錯乱枝を実験的に作成し, 能力群ごとに選択率の検討を行う. そこで, 本節では研究 2 への橋渡しとして, 先行研究の知見と本研究の結果を統合しながら, 研究 2 で検討対象とする錯乱枝について整理する.

本研究のカテゴリのうち, 5 番目以降のカテゴリは誤答選択枝に反映させる誤答として適切でないと判断されたため, 研究 2 では上位 4 個のカテゴリを取り上げることにした. 先行研究の知見を踏まえ, 特徴の類似性等の観点から, 錯乱枝となりうる誤答選択枝の条件を 3 要因 8 水準にまとめた (Table 2.2). 1 つ目の要因は「文章中の記述の有無」であり, その錯乱枝中に記述されている内容を誤りであると判断するための根拠が, 文章中に記述されているかどうかに関する要因であり, 「文章中の記述あり」「文章中の記述なし」の 2 水準から構成される. 文章中の記述あり錯乱枝は, Goodrich (1977) の文脈関連錯乱枝 (contextual distractor) に相当し, 文章中の記述なし錯乱枝は, Morimoto (2008) の無関連錯乱枝 (unrelated distractor) に相当する. 先行研究 (Goodrich, 1977; Morimoto, 2008) においては, 語彙能力を問う空所補充問題を材料に検討されていたが, 研究 2 においては英語文章読解問題を材料に, これらの錯乱枝の選択率に関する検討を行うこととする.

2 つ目の要因は「対概念の使用」であり, 本研究の誤答カテゴリである「対義語の使用」「対比関係の使用」の 2 カテゴリに加えて, Freedle & Kostin (1991) において, 誤答選択枝中に含まれる否定語の数が増加した場合に項目困難度が高くなるとの知見を考慮に入れ, 「否定語を使用した錯乱枝」を含めることとした. 最終的に, 「否定語の使用」「対義語の使用」「対比関係の使用」の 3 水準から「対概念の使用」要因を構成することとした. 否定

語の使用錯乱枝（以下，否定語錯乱枝）は，英語文章中の記述に否定語を追加して記述の意味を反転させたものである．誤答選択枝に否定語を含めることで，英語文章読解能力の低い受検者が否定語を含む誤答選択枝を選択しやすくなり，結果として正答選択枝の選択率が低くなった結果，項目困難度が高くなるものと考えられる．対義語の使用錯乱枝（以下，対義語錯乱枝）は，Goodrich (1977) の対義語錯乱枝に相当し，文章中の記述に含まれる語の対義語を誤答選択枝中に含めたものである．対比関係の使用錯乱枝（以下，対比関係錯乱枝）は，本研究における同名の誤答カテゴリに由来するものであり，文章全体の対比構造を利用した誤答選択枝である．対比関係錯乱枝は，英語文章読解ならではの錯乱枝であると言えるだろう．

3つ目の要因は「言及対象の取り違い」であり，本研究における同名の誤答カテゴリを取り上げたものである．上述のように，言及対象の取り違いに関する要因は，「内容の取り違い」「因果関係の取り違い」「比較対象の取り違い」の3水準から構成される．この要因における3種類の誤答は，King et al. (2004) および Lin et al. (2010) における錯乱枝の理論的分類の中で，Level 2「文章中に記述されている事実間の関係を取り違えている錯乱枝」に対応する．加えて，それぞれの誤答は，研究1における研究協力者のコメントを反映させたものである．

研究2では，以上の3要因8水準をさまざまに組み合わせ，多枝選択式の英語文章読解問題における錯乱枝の選択率に関する詳細な検討を行う．

## 2.3 研究2：受検者の能力群ごとにみた錯乱枝の選択率

研究2では，研究1で整理した錯乱枝の条件(3要因8水準)を組み合わせ，現実のテスト項目で想定されるさまざまな状況下での錯乱枝の選択率について，受検者の能力群ごとに検討する．

### 2.3.1 方法

#### 2.3.1.1 受検者

愛知県内の国立・私立大学に通う学生366名(男性92名，女性273名，不明1名)を受検者とした．研究1と同様，受検者は全員，英語以外を母語とする者であった．

#### 2.3.1.2 材料

研究1と同様，インターネット上で公開されている私立大学の入学試験問題の英語テストのうち，多枝選択式の文章読解問題を9題使用した(Table 2.3)．使用した9題のうち，1題は後述する共通項目用の英語文章読解問題として位置づけ，入学試験問題で出題されたまま使用したが，残り8題については実験項目用の英語文章読解問題として位置づけ，

Table 2.3  
研究 2 で使用した文章・設問一覧

冊子	テーマ	リーダビリティ	語数
共通	想像上の観衆*	14.64	284
冊子 1	ノーベル賞	7.01	411
冊子 2	著作権	10.79	519
冊子 3	無声映画	12.16	537
冊子 4	ルネサンス期の芸術家	8.92	498
冊子 5	ガムを噛むこと	7.85	580
冊子 6	研究*	10.39	455
冊子 7	乳児期における他者とのかかわり*	11.12	491
冊子 8	卵のしくみ	10.39	588

注) \*印は、研究 1 でも使用した文章であることを表す。

本研究の目的に合うよう作成した。実験項目用の英語文章読解問題の作成の詳細については、次節で述べることとする。

### 2.3.1.3 設問・選択枝の作成とテスト冊子の構成

研究 2 では、計 8 種類のテスト冊子を作成した (Table 2.4)。各テスト冊子は、第 1 問・第 2 問の大問 2 問構成であり、第 1 問では英語文章 1 題と多枝選択式問題 3 項目を、第 2 問では英語文章 1 題と多枝選択式問題 5 項目を提示した。なお、第 1 問と第 2 問の英語文章は、互いに異なるものになるようにテスト冊子を作成した。

各テスト冊子の大問 2 問のうち、第 1 問はすべての種類のテスト冊子を通じて同一の英語文章・多枝選択式問題（共通項目）を収録したが、第 2 問では、8 題の英語文章読解問題のうち 1 題を 1 種類のテスト冊子に収録した。このように、異なるテスト冊子間で共通項目を含む状況は、共通項目デザイン (common item design) と呼ばれている。共通項目デザインに基づいて項目を配置することにより、以降の分析で項目応答理論 (Item Response Theory; IRT) による等化が可能となり、受検者によって解答するテスト冊子が異なっても、受検者の潜在特性値を共通の尺度上で議論できるようになる。本論文の以降の章においても、多枝選択式問題に対する解答データを収集している場合には項目応答理論に基づく等化を念頭に入れ、共通項目デザインを採用している。

第 2 問では、本研究の目的に合わせて設問・選択枝を作り分けた実験項目とし、研究 1 において整理を行った 3 要因に対応させて、3 つの設問種別を設定した。それぞれの設問種別では、Table 2.2 を参照しながら、各要因に含まれる誤答の特徴と錯乱枝を 1 対 1 対応させた。3 つの設問種別のうち、文章中の記述の有無要因について検討を行うための項目として、正答選択枝の他に、文章中の記述あり錯乱枝・文章中の記述なし錯乱枝を作成したものを設問種別 A とした。また、対概念の使用要因について検討を行うための項目とし

問3 In the poem quoted in the text, Michelangelo wrote of his desire to 6.

- ① tell others what sufferings he had through his art (記述あり)
- ② fully realize himself through his art (正答)
- ③ make bigger and more beautiful works of art (記述なし)

問4 According to the third and fourth paragraph, Michelangelo 7.

- ① did not wonder why he wrote annoying letter when he was 22 years old (否定語)
- ② felt himself stuck in the psychological paradox of self reflection (正答)
- ③ thought that being sane was the only way to free from his distress (対義語)
- ④ was an exception of Renaissance artists, who were called a mad man (対比)

問5 Which of the following statements is true? 8

- ① Artists in the Renaissance spent their time working depending on the controlled routine.  
(内容の取り違い)
- ② Like Michelangelo, Goya talked to himself to persuade spiritually. (比較対象の取り違い)
- ③ Michelangelo forbade his friends to get in his atelier unless he wanted to be alone.  
(因果関係の取り違い)
- ④ Renaissance artists were satisfied to be involved in introspection. (正答)

Figure 2.3. 研究2で使用した設問の例

て、否定語錯乱枝・対義語錯乱枝・対比関係錯乱枝を作成したものを設問種別Bとした。さらに、言及対象の取り違い要因について検討を行うための項目として、内容の取り違い錯乱枝・因果関係の取り違い錯乱枝・比較対象の取り違い錯乱枝を作成したものを設問種別Cとした。なお、設問種別B・Cでは、複数の要因が組み合わさった状況下での錯乱枝の選択率について検討を行うため、以下に示したような詳細な状況設定を行った。

設問種別Bでは、文章中の記述の有無に関する実験操作を行った3つの状況下で、否定語錯乱枝・対義語錯乱枝・対比関係錯乱枝を提示した。具体的には、文章中の記述の有無を考慮しない状況(B1)、各選択枝の記述を誤りと判断するための根拠が文章中に存在する(文章中の記述あり)状況(B2)、選択枝中の記述を誤りと判断するための根拠が文章中に存在しない(文章中の記述なし)状況(B3)の3つの状況を設定した。

また、設問種別Cでは、文章中の記述の有無、対概念の使用をそれぞれ実験操作した6つの状況下で、内容の取り違い錯乱枝・因果関係の取り違い錯乱枝・比較対象の取り違

Table 2.4  
テスト冊子の構成

冊子 No.	第1問 (3項目)			第2問 (5項目)									
	共通項目			記述	対概念	言及	記述有 ・対概念	記述無 ・対概念	否定 ・言及	対義 ・言及	対比 ・言及	記述有 ・言及	記述無 ・言及
	A	B1	C1	B2	B3	C4	C5	C6	C2	C3			
冊子 1			問 1	問 2	問 3	問 4	問 5						
冊子 2						問 1	問 2	問 3	問 4	問 5			
冊子 3								問 1	問 2	問 3	問 4	問 5	
冊子 4	問 1	問 2	問 3	問 3	問 4	問 5						問 2	問 1
冊子 5				問 1	問 3	問 2	問 4	問 5					
冊子 6							問 2	問 1	問 3	問 4	問 5		
冊子 7									問 1	問 2	問 4	問 5	問 3
冊子 8				問 3	問 4	問 5						問 1	問 2

え錯乱枝を提示した。具体的には、文章中の記述の有無や対概念の使用を考慮しない状況 (C1)、文章中の記述あり状況 (C2)、文章中の記述なし状況 (C3)、錯乱枝中に否定語も含まれる状況 (C4)、錯乱枝中对義語が含まれる状況 (C5)、錯乱枝中に文章中の対比構造を用いた表現が含まれる状況 (C6) を設定した。なお、C4・C5・C6 の状況の設問では、すべての錯乱枝において否定語や対義語、文章中の対比構造を利用した表現を含むことになるため、受検者がテスト項目を見た際に不自然さを感じる可能性がある。本研究では、これらの設問種別において受検者が不自然さを感じないようにするため、同一項目で提示する錯乱枝中で使用語が重複しないよう留意して作成した。例えば C4 の設問では、not だけでなく without や few などの否定的な意味をもつ異なる語を複数使用した。

Table 2.4 に、各テスト冊子と設問種別・状況の対応表を示した。実験項目を配置した第2問では、各テスト冊子に5項目の実験項目を配置し、英語文章8題分で計40項目を作成した。その際、同一の設問種別に4項目ずつが割り当てられるよう配置し、それらが互いに異なるテスト冊子に含まれるようにした。

#### 2.3.1.4 手続き

解答データの収集は2013年6月～9月の間に、集団実施の形式で行われた。受検者に対し、本研究の目的を伝えた上で、8種類のテスト冊子のいずれか1種類をランダムに割り当てた。受検者にテスト冊子を配布した後、解答データが統計的に処理されるため個人を特定されないこと、テストの採点結果が受検者の成績評価とは関係しないこと、研究参加は任意であり、いつでも研究参加を中断できることなどをアナウンスした上で、受検者の研究参加への同意を確認した。

同意確認を行った後、受検者に解答にあたっての注意事項を教示した。具体的には、英語文章を読み、問いに対する答えとして正しいと思う選択枝の番号の丸を塗りつぶすよう

にして解答するよう伝えた。テストへの解答時間は25分間であった。

## 2.3.2 結 果

### 2.3.2.1 項目分析

受検者から得られた解答データに基づいて、項目分析を行った (Table 2.5)。項目分析を行うにあたり、第1問は入学試験問題のまま使用したため、項目ごとに分析を行ったが、第2問は誤答選択枝を本研究独自に作成したため、選択枝ごとに分析を行った。

本研究では項目分析の統計的指標として、困難度・識別力を用いた。困難度の指標として、第1問では正答率を、第2問では各選択枝の選択率を算出した。また、識別力の指標として、第1問では各項目への正誤 (正答=1, 誤答=0) と受検者の潜在特性値の EAP 推定値 (2.3.2.3 節で詳述) の点双列相関係数を、第2問では各選択枝への選択状況 (選択=1, 非選択=0) と受検者の潜在特性値の EAP 推定値との点双列相関係数を算出した。通常、識別力の指標として点双列相関係数を算出する場合には、各項目への正誤・各選択枝への選択状況と、その項目における得点を削除した正答数得点との相関係数を求めるが、本研究では受検者によって割り当てのあったテスト冊子が異なっており、正答数得点の解釈がテスト冊子によって異なることから、点双列相関係数として各項目への正誤・各選択枝への選択状況と受検者の潜在特性値の推定値の相関係数を算出した。

まず、正答選択枝に関する統計的指標に着目する。正答率および正答選択枝の選択率は0.1~0.9程度が望ましく、極端に正答率の高いもの・低いものは望ましくない (野口・大隅, 2014)。また、正答選択枝に関する点双列相関係数の値の慣習的な目安は0.2とされている (野口・大隅, 2014)。Table 2.5 の正答率・正答選択枝への選択率を見ると、困難度の高い (難しい) 項目で0.2程度、困難度の低い (易しい) 項目で0.8程度であった。8種類のテスト冊子全体で見ると、困難度は適正な範囲の中に収まっており、困難度にある程度のバリエーションがあったことがわかる。点双列相関係数に関しては、本研究において受検者に解答を求めた全43項目のうち、1項目 (冊子2・第2問・問5) において負の値を示しているものが見られ、2項目 (冊子1・第2問・問1, 冊子3・第2問・問2) において0.2を下回るものが見られた。

次に、誤答選択枝に関する統計的指標に着目する。誤答選択枝の選択率は、少なくとも5% (.05) 以上あることが望ましい (DiBattista & Kurzawa, 2011)。また、誤答選択枝に関する点双列相関係数は負の値を示すことが望ましいとされる (DiBattista & Kurzawa, 2011)。Table 2.5 において、誤答選択枝の選択率 (第2問) を見ると、おおむね5%以上の選択率となっているが、一部の選択枝において選択率が5% (.05) を下回るものがあり、誤答選択枝に関する点双列相関係数が正の値を示していたものもあった。

実際のテストにおいて項目分析を行ったとき、正答選択枝に関する点双列相関係数が

Table 2.5  
項目分析の結果

第1問	項目	正答率	点双列相関						
共通項目 (366名)	問1	.493	.498						
	問2	.288	.276						
	問3	.536	.451						
第2問	項目	選択枝	選択率	点双列相関	項目	選択枝	選択率	点双列相関	
冊子1 (44名)	問1	1(あり)	.500	-.125	冊子3 (44名)	問1	1(正答)	.163	.386
		2(正答)	.386	.163			2(内容)	.009	-.239
		3(なし)	.114	-.053			3(因果)	.581	-.191
	1(対義)	.114	-.289	4(比較)			.163	.057	
	問2	2(否定)	.091	-.254		問2	1(正答)	.209	.033
		3(正答)	.750	.494			2(内容)	.465	.032
		4(対比)	.045	-.237			3(因果)	.070	-.187
	1(内容)	.364	-.394	4(比較)			.256	.042	
	問3	2(因果)	.182	-.381		問3	1(比較)	.116	-.186
		3(比較)	.182	.109			2(正答)	.581	.516
		4(正答)	.273	.661			3(因果)	.209	-.395
	1(正答)	.386	.492	4(内容)			.093	-.118	
問4	2(否定)	.227	-.313	問4	1(内容)	.070	-.400		
	3(対比)	.318	-.250		2(因果)	.256	-.031		
	4(対義)	.068	.032		3(比較)	.186	-.091		
1(対比)	.250	-.320	4(正答)		.488	.302			
問5	2(否定)	.114	-.208	問5	1(内容)	.209	-.225		
	3(正答)	.455	.651		2(因果)	.023	-.389		
	4(対義)	.181	-.311		3(正答)	.558	.521		
1(正答)	.289	.231	4(比較)		.209	-.268			
冊子2 (45名)	問1	2(対比)	.133	-.081	冊子4 (46名)	問1	1(正答)	.522	.575
		3(否定)	.222	-.119			2(比較)	.196	-.276
		4(対義)	.356	-.058			3(内容)	.152	-.317
		1(対義)	.159	-.142			4(因果)	.130	-.189
	問2	2(正答)	.364	.586		問2	1(正答)	.378	.408
		3(対比)	.182	-.205			2(因果)	.222	.074
		4(否定)	.295	-.331			3(比較)	.289	-.367
	1(正答)	.341	.623	4(内容)			.111	-.198	
	問3	2(内容)	.227	-.371		問3	1(あり)	.457	-.385
		3(因果)	.205	-.175			2(正答)	.326	.567
		4(比較)	.227	-.165			3(なし)	.217	-.179
	1(内容)	.047	-.404	問4			1(否定)	.136	-.173
2(因果)	.163	-.438	2(正答)		.364	.574			
3(正答)	.651	.578	3(対義)		.341	-.300			
4(比較)	.140	-.083	4(対比)		.159	-.204			
問4	1(内容)	.302	.437	問5	1(内容)	.205	-.234		
	2(因果)	.233	-.014		2(比較)	.295	.008		
	3(正答)	.279	-.008		3(因果)	.205	-.087		
4(比較)	.186	-.493	4(正答)		.295	.275			

(次ページに続く)



Table 2.5  
項目分析の結果（続き）

第2問	項目	選択枝	選択率	点双列相関		項目	選択枝	選択率	点双列相関
冊子5 (47名)	問1	1(あり)	.362	-.381	冊子7 (48名)	問1	1(内容)	.208	-.260
		2(正答)	.362	.632			2(因果)	.208	-.290
		3(なし)	.277	-.269			3(正答)	.563	.483
			4(比較)	.021			-.116		
	問2	1(内容)	.064	-.156		問2	1(正答)	.354	.576
		2(比較)	.255	.024			2(内容)	.354	-.212
		3(正答)	.426	.281			3(因果)	.250	-.415
	4(因果)	.255	-.255	4(比較)			.042	.026	
	問3	1(対比)	.277	-.122		問3	1(比較)	.170	-.365
		2(対義)	.149	-.270			2(内容)	.170	-.191
		3(否定)	.191	-.408			3(正答)	.319	.506
	4(正答)	.383	.640	4(因果)			.340	-.057	
問4	1(対比)	.064	-.061	問4	1(内容)	.229	-.066		
	2(正答)	.553	.615		2(因果)	.271	-.497		
	3(否定)	.213	-.387		3(正答)	.438	.498		
4(対義)	.170	-.353	4(比較)		.063	.005			
問5	1(否定)	.239	-.092	問5	1(比較)	.354	-.097		
	2(対比)	.174	-.205		2(因果)	.167	-.004		
	3(正答)	.348	.521		3(内容)	.208	-.191		
4(対義)	.239	-.307	4(正答)		.271	.282			
冊子6 (48名)	問1	1(否定)	.250	-.151	冊子8 (44名)	問1	1(内容)	.182	-.147
		2(正答)	.479	.428			2(正答)	.227	.146
		3(対義)	.083	-.215			3(因果)	.545	-.034
		4(対比)	.188	-.229			4(比較)	.045	-.059
	問2	1(正答)	.234	.382		問2	1(因果)	.047	-.085
		2(否定)	.043	-.147			2(正答)	.674	.432
		3(対義)	.511	-.114			3(内容)	.116	-.213
	4(対比)	.213	-.182	4(比較)			.162	-.315	
	問3	1(内容)	.375	-.239		問3	1(正答)	.465	.368
		2(正答)	.208	.238			2(なし)	.209	-.481
		3(因果)	.167	-.091			3(あり)	.326	.026
	問4	4(比較)	.250	.122		問4	1(対義)	.143	-.291
1(正答)		.435	.500	2(否定)	.048		-.230		
2(内容)		.304	-.281	3(正答)	.738		.457		
3(因果)	.130	-.321	4(対比)	.071	-.195				
問5	4(比較)	.130	-.105	問5	1(内容)	.163	-.037		
	1(因果)	.222	-.224		2(比較)	.163	-.317		
	2(比較)	.111	-.248		3(因果)	.233	-.209		
3(正答)	.511	.500	4(正答)		.442	.440			
		4(内容)	.156	-.217					

0.2を下回る、誤答選択枝の選択率が5%を下回る、誤答選択枝に関する点双列相関係数が正の値を示すなどの特徴が見られた場合には、テスト項目の修正(e.g., 正答選択枝が真に正答になっているかどうかの確認、誤答選択枝の修正など)を行ったり、別の項目と取りかえることなどが必要になる。ただし、本研究ではテストを研究目的で実施しており、錯乱枝の特徴や機能について検討することを目的としているため、選択率・点双列相関係数

が低い項目についても以降の分析対象とした。

### 2.3.2.2 受検者の得点分布

受検者のテスト得点（正答数得点）のヒストグラムを Figure 2.4 に示した。受検者の得点の平均は 3.34 点、標準偏差は 1.70 点であった。

テスト冊子の種類に応じて受検者のテスト得点が変わるかどうか確認を行うため、テスト冊子ごとに得点の平均値と標準偏差を算出した (Table 2.6)。受検者の得点の平均値は、テスト冊子全体を通して 3.00 点～4.00 点の間にあり、テスト冊子の種類の違いに応じて受検者のテスト得点に差があるとは見なせなかった。

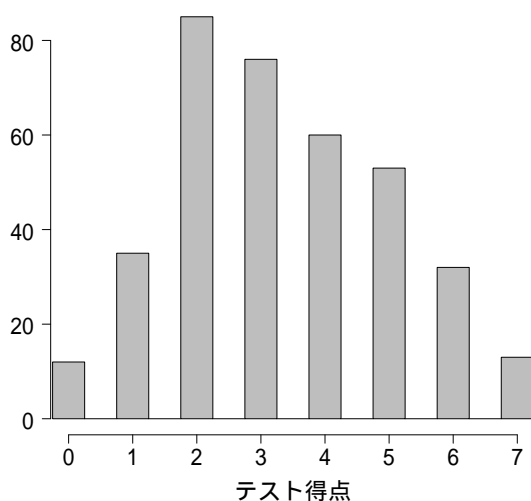


Table 2.6. 冊子ごとの得点の記述統計量

冊子	<i>M</i>	<i>SD</i>
冊子 1	3.45	1.84
冊子 2	3.16	1.88
冊子 3	3.18	1.42
冊子 4	3.22	1.86
冊子 5	3.43	1.81
冊子 6	2.96	1.47
冊子 7	3.31	1.64
冊子 8	4.02	1.56

Figure 2.4. 受検者の得点分布

### 2.3.2.3 受検者の潜在特性値の推定

本研究では上述の通り、項目応答理論を用いた潜在特性尺度の等化を念頭に入れ、共通項目デザインを適用して解答データを収集した。以下では、項目応答理論について簡潔に説明を行った上で、受検者の潜在特性値の推定結果について述べる。

本研究では、項目の困難度 ( $b_j$ )・識別力 ( $a_j$ ) の推定を主な関心とはせず、異なるテスト冊子に解答した受検者の潜在特性値の推定値を得ることが目的であったため、1パラメタ・ロジスティック・モデル (以下、1PLM) を採用した。1PLM の項目特性関数 (item characteristic function) は、以下の式で表現される。

$$p_j(\theta_i) = \frac{1}{1 + \exp[-Da(\theta_i - b_j)]} \quad (2.1)$$

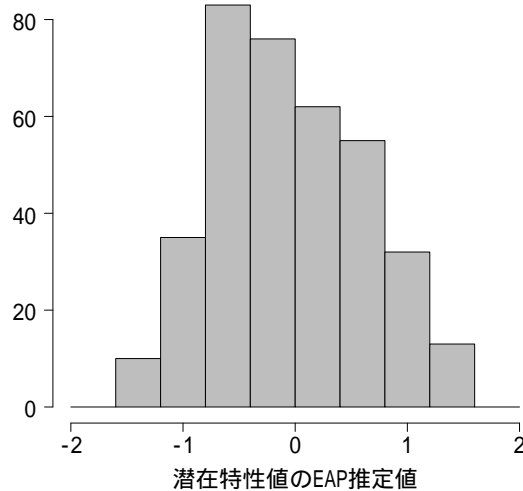


Figure 2.5. 受検者の潜在特性値の推定値のヒストグラム

ここで、左辺の  $p_j(\theta_i)$  は受検者  $i$  が項目  $j$  に正答する確率を表している。また、右辺の  $a$  は識別力、 $b_j$  は困難度、 $\theta_i$  は受検者  $i$  の潜在特性値を表している。本研究では、受検者の潜在特性値  $\theta_i$  を推定の上、受検者の能力値とした。なお、定数  $D$  の設定に関して、正規累積モデルへの近似が必要となる場合には  $D = 1.7$  とするが、近似の必要性がない場合には  $D = 1.0$  とするのが一般的である。本研究では、正規累積モデルへの近似が特に必要ではないと判断し、 $D = 1.0$  とした。また、識別力  $a$  に関しては、すべての項目において同一となるように推定される。

項目パラメタおよび受検者の潜在特性値の推定に用いたデータ行列は、受検者 366 名  $\times$  43 項目 (共通項目 3 項目 + 1 種類のテスト冊子につき実験項目 5 項目  $\times$  8 種類) から構成される。本研究では、8 種類のテスト冊子のうち 1 種類を受検者にランダムに割り当てて解答を求めたため、例えば冊子 1 に解答した受検者は、冊子 2~冊子 8 の問題には解答していないことになり、冊子 2~冊子 8 のデータが欠測となる。このように、受検者が解答しなかったテスト冊子に含まれる項目への応答がすべて欠測値となった行列であっても、共通項目が含まれている場合には、同時尺度調整法 (concurrent calibration) によって、項目パラメタ・受検者の潜在特性値を推定することが可能である。本研究では、共通項目をアンカーとして同時尺度調整法による等化を行い、異なるテスト冊子に解答した受検者の潜在特性値を共通尺度上に位置づけた。

パラメタの推定にあたり、項目の困難度  $b_j$  および識別力  $a$  に関しては周辺最尤推定法を適用した。また、潜在特性値  $\theta_i$  に関しては、ベイズ推定法のうち事後期待値 (Expectation A Posteriori; EAP) に基づく推定値を用いた。なお、受検者母集団における潜在特性値  $\theta$  の

事前分布として、慣習的によく用いられる標準正規分布  $N(0,1)$  を設定した。

Figure 2.5 に、受検者の潜在特性値の EAP 推定値のヒストグラムを示した。受検者の潜在特性値の推定値の平均は  $-0.04$ 、標準偏差は  $0.65$  であった。

#### 2.3.2.4 受検者の能力群の構成

受検者の潜在特性値の EAP 推定値を用いて、受検者の能力群を構成した。一般に、量的変数を用いて群分けを行うことは、データがもつ情報量を過度に削ぎ落とすことになるため、量的変数は量的変数のまま扱うことが望ましいとされている。一方、項目分析において選択枝の選択状況に関する検討を行う際、受検者群と選択枝の選択状況について集計を行った分割表や、受検者群ごとの選択率を折れ線で示したトレースラインがよく用いられる (Haladyna & Rodriguez, 2013)。これらの分析に共通する点は、その項目における選択枝への選択状況を解釈しやすくするために、受検者の能力値に基づいて群分けを行っているということである。本研究でも、錯乱枝への選択状況に対する解釈の容易さを重視して、受検者の能力群を用いた分析を行った。

受検者の能力群の構成に際して、本研究では受検者の潜在特性値の推定値を能力値とし、項目分析において伝統的に用いられている Kelley (1939) の方法を適用して、受検者の能力分布の 27 パーセンタイル点以下の受検者を能力低群、27 パーセンタイル点から 73 パーセンタイル点までの受検者を能力中群、73 パーセンタイル以上の受検者を能力高群として、受検者を 3 群に分割した。Kelley (1939) の方法は、能力低群と高群の選択率の差を最大化しながら、各能力群の人数が少なくなりすぎない臨界点を算出したものであるとされており (Cureton, 1957; 池田, 1973)、テストの項目分析を行った実証的研究において現在もなお用いられている方法である (e.g., 安永・石井, 2012)。

Kelley (1939) の方法に沿って、受検者の潜在特性値の推定値の分布の 27 パーセンタイル点 ( $q(.27) = -0.541$ ) と 73 パーセンタイル点 ( $q(.73) = 0.460$ ) を求めた上で、受検者を 3 群に分割した。その結果、能力低群 104 名 ( $M = -0.796$ ,  $SD = 0.263$ )、能力中群 163 名 ( $M = -0.077$ ,  $SD = 0.259$ )、能力高群 99 名 ( $M = 0.806$ ,  $SD = 0.275$ ) に分割された。

#### 2.3.2.5 能力群ごとにみた錯乱枝の選択率に関する検討

設問種別ごとに各能力群の錯乱枝の選択率を把握するため、トレースラインを描出した (Figure 2.6~2.8)。トレースラインは、設問解答率分析図とも呼ばれ、受検者の能力段階を横軸に、各選択枝の選択率を縦軸にとった折れ線グラフのことである。この図は、各選択枝の選択状況を受検者の能力群ごとに集計した分割表をもとに描出されるものである。

トレースラインは一般に、各選択枝の選択状況を項目単位で把握するためのツールとしてよく用いられるが、本論文では以降の章を含めて、項目単位ではなく、設問種別や条件、認知過程単位などの単位でトレースラインを描出し、同一の設問種別・条件・認知過程に

割り当てられた複数の項目についての結果を集約した形で表示した。もちろん、英語文章や項目の違いに応じて各選択枝の選択率は異なるものであるが、本論文では、英語文章・項目の違いを含み込んで、能力群間で選択率が異なるかどうかといった点や、設問種別や条件などの違いによって選択率が異なるかどうかといった点を明らかにすることが主眼にあるため、項目単位の分析ではなく、複数の項目の結果を集約したそれぞれの単位での分析を行った。

以下では、設問種別 A に対する分析結果を分析 1 として、設問種別 B1~B3 に対する分析結果を分析 2 として、設問種別 C1~C6 に対する分析結果を分析 3 として記載した。また、本章が本論文において錯乱枝の分析結果を記述する最初の章である点を考慮し、分析対象とした分割表についての記載や、判断に用いた統計量についての数理的な説明を簡単に行った。

#### 分析 1：文章中の記述の有無に関する錯乱枝の選択率の分析

**分析 1 が対象とする分割表とオッズ比の推定** 本分析の対象は、受検者の能力群と、設問種別 A に割り当てられたテスト項目 4 問に対する応答を結合した上での各選択枝の選択状況を整理した、 $3 \times 3$  の 2 元分割表である (Table 2.7)。

Table 2.7  
分析 1 で扱う分割表

	正答	文章中の 記述あり	文章中の 記述なし
能力高群	$p_{H1}$	$p_{H2}$	$p_{H3}$
能力中群	$p_{M1}$	$p_{M2}$	$p_{M3}$
能力低群	$p_{L1}$	$p_{L2}$	$p_{L3}$

Table 2.7 のセルには、能力群ごとにみた各選択枝の選択率 ( $p$ ) を記しており、選択率  $p$  の第 1 添え字は能力群を、第 2 添え字は選択枝の種類を表している。能力高群は  $H$ 、能力中群は  $M$ 、能力低群は  $L$  という添え字を付しており、正答選択枝の場合は 1、文章中の記述あり錯乱枝の場合は 2、文章中の記述なし錯乱枝の場合は 3 という添え字を付している。これらの選択率は、能力群内で合計すると 1 になるように計算したものである。

$$\sum_{k=1}^3 p_{gk} = 1, \quad g \in \{H, M, L\}. \quad (2.2)$$

本論文では、錯乱枝の選択状況に関する分析にあたり、能力群と選択枝の選択率について集計した分割表をもとにオッズ比を構成することとした。第 1 章でも述べたとおり、本論文で選択率を用いた理由として、各研究での各能力群の受検者数の違いに左右されず

に、一貫した解釈を行うことができる点が挙げられる。錯乱枝について分析を行う際、その錯乱枝の選択人数を用いる方法もあるが (Haladyna & Rodriguez, 2013), 受検者数が各能力群で等しいことが前提となっている。本論文で用いた方法は前節にも述べたとおり、能力低群と能力高群の選択率の差を最大化する群分けの方法であるため、結果として各能力群の受検者数に違いが生じる。こうした状況では、選択人数を分析対象とすることが適切でない可能性もある。一方で、選択率は各能力群に属する受検者数の違いに左右されずに「錯乱枝の選ばれやすさ」を表す指標であるため、複数の研究を通して一貫した解釈を行うことができるものと考えられる。以上の理由から、本論文では選択率に対して検討を行うこととした。

本論文の分析で扱う分割表は、行数  $r$ , 列数  $c$  の  $r \times c$  個のセルをもつ。Table 2.7 では、能力群を行に、選択枝を列に配置しており、能力群数は 3, 選択枝数も 3 であるため、 $3 \times 3$  の分割表ということになる。このような分割表に対するオッズ比のひとつとして、局所オッズ比 (local odds ratio) がある。局所オッズ比とは、 $r \times c$  の分割表における任意の 4 セルを取り出して構成するオッズ比のことである (Agresti, 2002; Fagerland et al., 2017)。 $r \times c$  の分割表において、局所オッズ比は  $(r-1)(c-1)$  個構成されるが、取り出す任意の 4 セルの組み合わせは唯一に定まるものではなく、検討したいことがらに応じて異なる (Fagerland et al., 2017)。

分析 1 では、正答選択枝の選択率に対する各錯乱枝の選択率のオッズを算出し、そのオッズを隣り合う能力群間 (能力低群—能力中群, 能力中群—能力高群) で比較するようなオッズ比を構成した。正答選択枝の選択率を基準としたのは、錯乱枝の選択率に対する解釈が正答選択枝の選択率の状況によって異なるという点を考慮するためである。実際、錯乱枝の選択率が同一の値をとっていた場合でも、正答選択枝の選択率が高い状況では「錯乱枝はそれほど選択されていない」と解釈し、正答選択枝の選択率が低い状況では「錯乱枝はよく選ばれている」と解釈することが、錯乱枝の機能を適切に反映するものと考えられる。このような理由から、オッズの構成にあたって正答選択枝の選択率を基準とした。

能力高群と能力中群の間で正答選択枝 ( $k = 1$ ) の選択率に対する錯乱枝  $k$  ( $k \neq 1$ ) の選択率のオッズを比較するための局所オッズ比は、

$$OR = \frac{p_{Mk}/p_{M1}}{p_{Hk}/p_{H1}} \quad (2.3)$$

として構成し、能力中群と能力低群の間で正答選択枝の選択率に対する錯乱枝  $k$  の選択率のオッズを比較するための局所オッズ比は、

$$OR = \frac{p_{Lk}/p_{L1}}{p_{Mk}/p_{M1}} \quad (2.4)$$

として構成した。

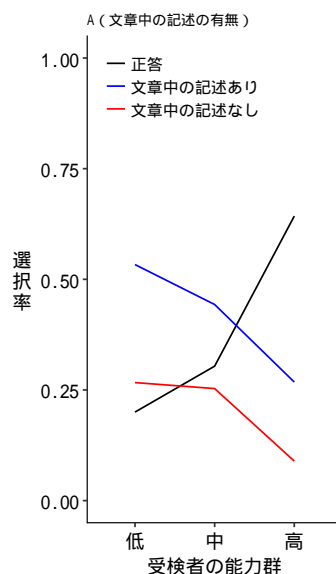


Figure 2.6. 文章中の記述の有無（設問種別 A）のトレースライン

ここで、能力低群と中群の間の「文章中の記述あり錯乱枝」の選択率について比較を行う場合を例に、局所オッズ比に対する解釈を整理したい。(2.4)式の分母・分子はいずれも、当該錯乱枝の選択率を正答選択枝の選択率で割った値となっている。この値は、能力低群・中群それぞれで文章中の記述あり錯乱枝の選択率が正答選択枝の選択率の何倍であったかを表しており、正答選択枝に対する錯乱枝の「選ばれやすさ」を反映していると考えられることも可能である。具体的な値を用いて考えると、能力低群では正答選択枝の選択率が0.200、文章中の記述あり錯乱枝の選択率が0.533であったことから、文章中の記述あり錯乱枝は正答選択枝に比べ、 $0.533/0.200 = 2.665$ 倍選ばれやすいということがわかる。同様に、能力中群では正答選択枝の選択率が0.304、文章中の記述あり錯乱枝の選択率が0.443であったことから、文章中の記述あり錯乱枝は正答選択枝に比べ、 $0.443/0.304 = 1.457$ 倍選ばれやすいということがわかる。

(2.4)式の局所オッズ比は、錯乱枝の「選ばれやすさ」について能力群間で比較することを意味する。能力中群よりも能力低群で文章中の記述あり錯乱枝が選ばれやすい場合、局所オッズ比は1より大きな値となる。実際、能力低群では中群の1.829倍(=  $2.665/1.457$ )選ばれやすいということが示唆される。

標本オッズ比  $\widehat{OR}$  の対数  $\log \widehat{OR}$  の標本分布は、能力群  $g$  で選択枝  $k$  を選んだ受検者数を

Table 2.8

分析 1 におけるオッズ比の推定結果

能力群	文章中の記述あり錯乱枝		文章中の記述なし錯乱枝	
	OR	95%CI	OR	95%CI
能力中群—能力高群	3.500	[1.580, 7.753]	6.000	[1.982, 18.165]
能力低群—能力中群	1.829	[0.725, 4.614]	1.600	[0.561, 4.565]

$n_{gk}$  としたとき、平均  $\log OR$ 、漸近分散が

$$\text{Asymp.Var}(\log \widehat{OR}) = \frac{1}{n_{g1}} + \frac{1}{n_{gk}} + \frac{1}{n_{g'1}} + \frac{1}{n_{g'k}}, \quad g \neq g', k \geq 2. \quad (2.5)$$

の正規分布に漸近的に従うことが知られている (Agresti, 2002). なお、この標本分布の分散は、デルタ法 (delta method) に基づいて Taylor 近似された漸近分散である。本論文では、対数オッズ比の漸近標準誤差 ASE (標本分布の漸近分散の正の平方根) を用い、対数オッズ比の 95% 信頼区間の下限を  $\log \widehat{OR} - 1.96\text{ASE}$ 、上限を  $\log \widehat{OR} + 1.96\text{ASE}$  として算出した上で、下限値と上限値の指数をとってオッズ比の漸近的な 95% 信頼区間を構成した。

**分析 1 におけるオッズ比の推定結果** 設問種別 A のトレースラインを Figure 2.6 に、分析 1 におけるオッズ比の推定結果を Table 2.8 に示した。文章中の記述あり錯乱枝・文章中の記述なし錯乱枝ともに、中群—高群間では大きなオッズ比が得られ、中群の選択率が高群の選択率に比べて高いことが示されたが、低群—中群間ではオッズ比が 1 に近い値であり、低群—中群間で選択率の顕著な差は見られなかった。

## 分析 2：対概念の使用に関する錯乱枝の選択率の分析

**分析 2 が対象とする分割表とオッズ比の推定** 分析 2 が検討対象とする分割表は、受検者の能力群 (3)、設問種別 (3)、各選択枝の選択状況 (4) を整理した、 $3 \times 3 \times 4$  の 3 元分割表である (Table 2.9, 2.10)。3 変数以上の分割表に対して局所オッズ比を求める場合には、いずれか 1 つの変数を所与として 2 変数間の分割表を作成した上で、局所オッズ比を算出することになる。

分析 2 では、同一の能力群内における設問種別間の選択率の違いに関する検討と、同一の設問種別における能力群間の選択率の違いに関する検討の両方を行うため、受検者の能力群を条件づけた場合における設問種別間のオッズ比の検討 (能力群内の検討) を分析 2-1 とし、設問種別を条件づけた場合における能力群間のオッズ比の検討 (能力群間の検討) を分析 2-2 として、条件つき局所オッズ比に関する推測を行った。なお、信頼区間の構成には、分析 1 と同様、漸近分布を用いる方法を採用した。

分析 2-1 では、能力群ごとに設問種別  $\times$  選択枝の分割表を作成し、設問種別 B1 における正答選択枝の選択率を基準として、能力群内で設問種別間の選択率の違いを検討するためのオッズ比を構成した。例えば、能力高群における否定語錯乱枝の選択率を、設問種別



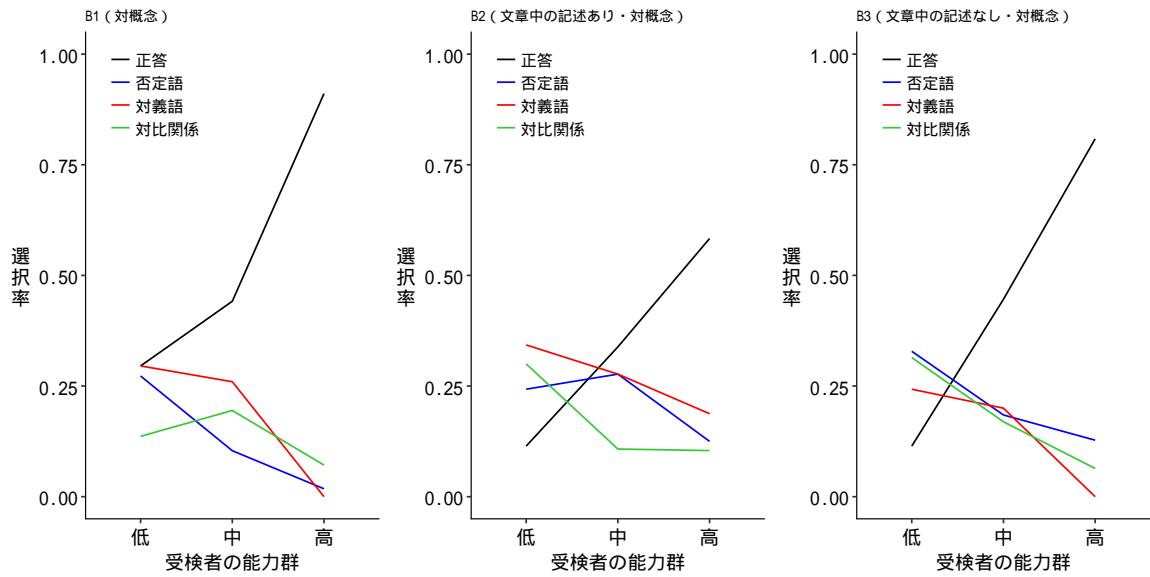


Figure 2.7. 対概念の使用（設問種別 B）のトレースライン

B1 と B2 で比較するためのオッズ比は、

$$OR = \frac{P_{B2,2|H}/P_{B2,1|H}}{P_{B1,2|H}/P_{B1,1|H}} \quad (2.6)$$

により構成した。

一方、分析 2-2 では、設問種別ごとに能力群 × 選択枝の分割表を作成し、設問種別  $q$  ( $q \in \{B1, B2, B3\}$ ) において、能力群間の選択率に関する検討を行うためのオッズ比を構成した。例えば、設問種別 B1（実験操作なし）の否定語錯乱枝について、能力高群と能力中群のオッズの大きさを比較するためのオッズ比は、

$$OR = \frac{P_{M2|B1}/P_{M1|B1}}{P_{H2|B1}/P_{H1|B1}} \quad (2.7)$$

により構成した。

分析 2-1 におけるオッズ比の推定結果 設問種別 B1~B3 のトレースラインを Figure 2.7 に、分析 2-1 におけるオッズ比の推定結果を Table 2.11 に示した。否定語錯乱枝に関し

Table 2.9

分析 2-1 で扱う分割表

能力群 $g$ ( $g \in \{H, M, L\}$ )	設問種別 $q$ ( $q \in \{B1, B2, B3\}$ )			
	正答	否定語	対義語	対比関係
設問種別 B1	$P_{b1 g}$	$P_{b2 g}$	$P_{b3 g}$	$P_{b4 g}$
設問種別 B2	$P_{d1 g}$	$P_{d2 g}$	$P_{d3 g}$	$P_{d4 g}$
設問種別 B3	$P_{e1 g}$	$P_{e2 g}$	$P_{e3 g}$	$P_{e4 g}$

Table 2.10

分析 2-2 で扱う分割表

設問種別 $q$ ( $q \in \{B1, B2, B3\}$ )	能力群 $g$ ( $g \in \{H, M, L\}$ )			
	正答	否定語	対義語	対比関係
能力高群	$P_{H1 q}$	$P_{H2 q}$	$P_{H3 q}$	$P_{H4 q}$
能力中群	$P_{M1 q}$	$P_{M2 q}$	$P_{M3 q}$	$P_{M4 q}$
能力低群	$P_{L1 q}$	$P_{L2 q}$	$P_{L3 q}$	$P_{L4 q}$

Table 2.11

## 分析 2-1 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		対比関係錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力高群</b>						
文章中の記述あり	7.830	[1.252, 48.980]	34.333	[1.923, 611.864]	2.209	[0.586, 8.327]
文章中の記述なし	5.797	[0.936, 35.912]	1.338	[0.026, 68.926]	1.040	[0.242, 4.471]
<b>能力中群</b>						
文章中の記述あり	3.337	[1.265, 8.807]	1.384	[0.607, 3.151]	0.742	[0.267, 2.059]
文章中の記述なし	1.719	[0.632, 4.677]	0.770	[0.331, 1.792]	0.868	[0.350, 2.150]
<b>能力低群</b>						
文章中の記述あり	2.224	[0.723, 6.842]	2.882	[0.974, 8.533]	5.253	[1.540, 17.924]
文章中の記述なし	2.986	[0.995, 8.959]	2.059	[0.676, 6.271]	5.498	[1.617, 18.697]

注) 実験操作なし条件における正答選択枝の選択率を参照セルとした。

ては、能力中群・能力高群の2群において、文章中の記述あり条件でオッズ比が大きいという結果が得られた。この結果から、能力中群・高群においては、すべての選択枝について正誤判断に必要な情報が英語文章中に存在するよう作成された設問で、否定語錯乱枝の選択率が高くなることが示唆された。また、対義語錯乱枝に関しては、能力高群においてのみ文章中の記述あり条件のオッズ比が大きくなっていった。このような結果が得られた背景には、実験操作なし条件・文章中の記述なし条件において能力高群の対義語錯乱枝の選択率が極めて低かったのに対し、文章中の記述あり条件における能力高群の対義語錯乱枝の選択率のみ高かったことが挙げられる。一方、対比関係錯乱枝に関しては、能力低群のみにおいて、文章中の記述あり条件・文章中の記述なし条件ともにオッズ比が大きく、他の能力群では設問種別間の選択率の違いは検出されなかった。

**分析 2-2 におけるオッズ比の推定結果** 分析 2-2 におけるオッズ比の推定結果を Table 2.12 に示した。否定語錯乱枝に関して、中群—高群の間では、実験操作なし条件・文章中の記述あり条件でオッズ比が大きかったが、低群—中群の間では、実験操作なし条件・文章中の記述なし条件で大きなオッズ比が得られた。これらの結果から、否定語錯乱枝は、必要な情報が英語文章中に含まれている設問（文章中の記述あり条件）では能力中群と高群の受検者を識別するのに有用であるが、英語文章中に記述されていない内容を含む設問（文章中の記述なし条件）では能力低群と中群の受検者を識別するのに有用であることが示唆された。

対義語錯乱枝に関して、能力中群—高群の間では、実験操作なし条件・文章中の記述なし条件において大きなオッズ比が得られたが、能力低群—中群の間では、文章中の記述あり条件・文章中の記述なし条件において大きなオッズ比が得られた。能力中群—高群間における対義語錯乱枝のオッズ比が大きくなっていった背景には、実験操作なし条件・文章中の記述なし条件において、能力高群のうち対義語錯乱枝を選ぶ者が極めて少なかったこと

Table 2.12  
分析 2-2 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		対比関係錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力中群—能力高群</b>						
実験操作なし	8.459	[1.414, 50.590]	61.203	[3.582, 1045.810]	5.142	[1.653, 15.991]
文章中の記述あり	3.605	[1.261, 10.310]	2.467	[0.946, 6.429]	1.727	[0.504, 5.917]
文章中の記述なし	2.510	[0.868, 7.260]	35.237	[2.012, 617.221]	4.288	[1.182, 15.563]
<b>能力低群—能力中群</b>						
実験操作なし	3.758	[1.283, 11.009]	1.683	[0.663, 4.272]	1.072	[0.353, 3.256]
文章中の記述あり	2.504	[0.899, 6.976]	3.506	[1.298, 9.468]	7.588	[2.412, 23.872]
文章中の記述なし	6.525	[2.338, 18.205]	4.499	[1.587, 12.757]	6.790	[2.394, 19.260]

が挙げられる。これらの結果から、文章中の記述なし条件における対義語錯乱枝は、各能力群間の識別に有効であった一方で、実験操作なし条件下での対義語錯乱枝は中群—高群間の識別に、文章中の記述あり条件下での対義語錯乱枝は低群—中群間の識別に有効であることが示された。

対比関係錯乱枝についても対義語錯乱枝と同様の傾向が見い出され、能力中群—高群の間では実験操作なし条件・文章中の記述なし条件においてオッズ比が大きくなっており、能力低群—中群の間では文章中の記述あり条件・なし条件で大きなオッズ比が得られた。能力中群—高群間の選択率の差は対義語錯乱枝ほどではなかったものの、約4倍～5倍ほどのオッズ比は得られていた。この結果から、実験操作なし条件・文章中の記述あり条件では、能力中群の選択率が能力高群よりも高かったことが示唆される。また、能力低群—中群では、文章中の記述あり条件・なし条件ともに対比関係錯乱枝に関するオッズ比が大きかったことから、文章中の記述あり条件・なし条件ともに、能力低群の選択率が中群よりも高かったことがうかがえる。

### 分析 3：言及対象の取り違えに関する錯乱枝の選択率の分析

**分析 3 が対象とする分割表とオッズ比の推定** 分析 3 での検討対象は、受検者の能力群 (3)、設問種別 (6)、各選択枝の選択状況 (4) を整理した、 $3 \times 6 \times 4$  の 3 元分割表である。以下では分析 2 と同様、受検者の能力群を条件づけた場合における設問種別間のオッズ比の検討（能力群内の検討）を分析 3-1 とし、設問種別を条件づけた場合における能力群間のオッズ比の検討（能力群間の検討）を分析 3-2 として、条件つき局所オッズ比に関する推測を行った。なお、信頼区間の構成には、分析 1・2 と同様、漸近分布を用いる方法を採用した。

**分析 3-1 におけるオッズ比の推定結果** 設問種別 C1～C6 のトレースラインを Figure 2.8 に、分析 3-1 におけるオッズ比の推定結果を Table 2.13 に示した。能力群内において実験操作なし条件よりも選択率が高くなっていたのは、文章中の記述あり条件下での能力高群

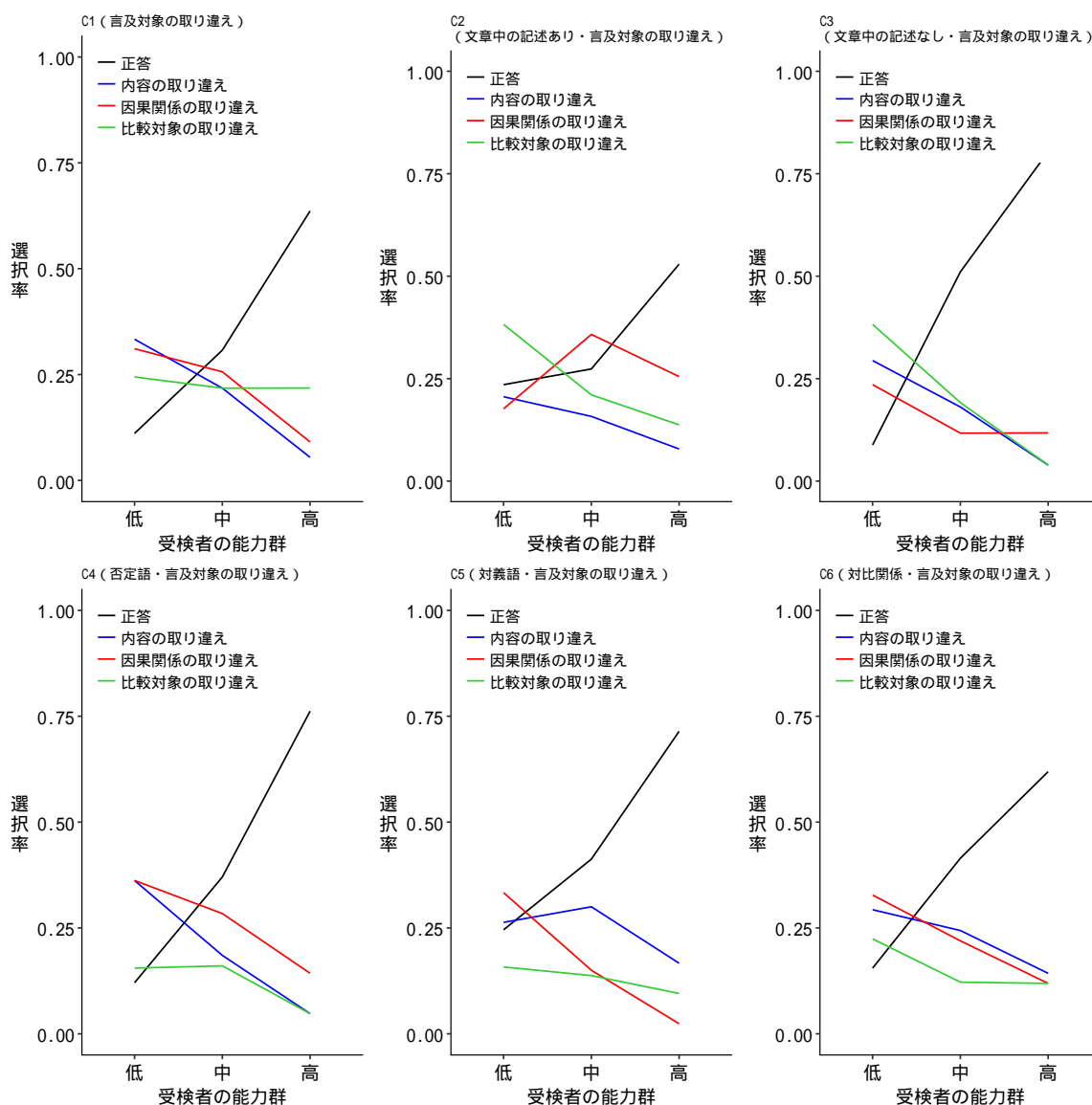


Figure 2.8. 言及対象の取り違え（設問種別 C）のトレースライン

における因果関係の取り違え錯乱枝のみであった。能力高群の受検者において、実験操作を行わなかった設問よりも、選択枝の正誤判断に必要な記述が英語文章中に存在する設問において、因果関係の取り違え錯乱枝の選択率が高かった。Figure 2.8 における設問種別 C3 のトレースラインを他の条件のものと比較しても、因果関係の取り違え錯乱枝の選択率が高いことが見てとれる。

Table 2.13  
分析 3-1 のオッズ比の推定結果

条件	内容の取り違い錯乱枝		因果関係の取り違い錯乱枝		比較対象の取り違い錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力高群</b>						
文章中の記述あり	1.728	[0.356, 8.382]	3.370	[1.070, 10.613]	0.756	[0.262, 2.180]
文章中の記述なし	0.569	[0.090, 3.602]	1.024	[0.288, 3.646]	0.142	[0.030, 0.679]
否定語	0.729	[0.114, 4.648]	1.313	[0.365, 4.720]	0.182	[0.038, 0.878]
対義語	2.722	[0.646, 11.464]	0.233	[0.026, 2.109]	0.389	[0.133, 1.333]
対比関係	2.692	[0.615, 11.779]	1.346	[0.353, 5.138]	0.561	[0.176, 1.789]
<b>能力中群</b>						
文章中の記述あり	0.814	[0.335, 1.981]	1.569	[0.717, 3.433]	1.086	[0.463, 2.546]
文章中の記述なし	0.500	[0.218, 1.149]	0.275	[0.114, 0.666]	0.529	[0.232, 1.207]
否定語	0.706	[0.293, 1.698]	0.920	[0.412, 2.057]	0.612	[0.249, 1.504]
対義語	1.207	[0.455, 2.317]	0.436	[0.180, 1.061]	0.471	[0.187, 1.184]
対比関係	0.830	[0.362, 1.907]	0.635	[0.279, 1.488]	0.415	[0.162, 1.063]
<b>能力低群</b>						
文章中の記述あり	0.292	[0.070, 1.222]	0.268	[0.062, 1.166]	0.739	[0.187, 2.924]
文章中の記述なし	1.111	[0.216, 5.727]	0.952	[0.179, 5.081]	1.970	[0.382, 10.166]
否定語	1.000	[0.266, 3.763]	1.071	[0.283, 4.059]	0.584	[0.138, 2.483]
対義語	0.357	[0.103, 1.243]	0.485	[0.141, 1.663]	0.292	[0.076, 1.126]
対比関係	0.630	[0.172, 2.299]	0.754	[0.207, 2.747]	0.657	[0.169, 2.549]

注) 実験操作なし条件における正答選択枝の選択率を参照セルとした。

分析 3-1 では、実験操作なし条件よりも選択率が低くなっていた条件が観測された。具体的には、能力中群における文章中の記述なし条件の因果関係の取り違い錯乱枝、能力高群における文章中の記述なし条件・否定語条件の比較対象の取り違い錯乱枝で、実験操作なし条件よりも選択率が低いという結果が得られた。実際、Figure 2.8 を見ても、C3 における能力中群の因果関係の取り違い錯乱枝は、C1 における能力中群のものよりも選択率が低いことが見て取れる。同様に、C3 における能力高群の比較対象の取り違い錯乱枝は、C1 における能力高群の比較対象の取り違い錯乱枝よりも低くなっていた。また、C4 における能力高群の比較対象の取り違い錯乱枝も、C1 における能力高群の比較対象の取り違い錯乱枝に比べて選択率が低いことが確認できる。

**分析 3-2 におけるオッズ比の推定結果** 分析 3-2 におけるオッズ比の推定結果を Table 2.14 に示した。内容の取り違い錯乱枝に関して、おおむね能力群間の差が見られたが、対比関係条件において能力中群と能力高群の間のオッズの違いは見られず、文章中の記述あり条件・対義語条件において能力低群と能力中群の間のオッズの違いが見られなかった。トレースラインを確認すると、上記の条件においては内容の取り違い錯乱枝の選択率が能力群間でさほど異ならなかったことから、能力群間の差を浮き彫りにするような錯乱枝でなかったことが見てとれる。

Table 2.14

## 分析 3-2 におけるオッズ比の推定結果

条件	内容の取り違え錯乱枝		因果関係の取り違え錯乱枝		比較対象の取り違え錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力中群—能力高群</b>						
実験操作なし	8.264	[2.180, 31.333]	5.833	[1.924, 17.685]	2.066	[0.837, 5.098]
文章中の記述あり	3.894	[1.141, 13.288]	2.716	[1.924, 17.685]	2.967	[1.075, 8.190]
文章中の記述なし	7.260	[1.583, 33.306]	1.566	[0.533, 4.604]	7.688	[1.683, 35.120]
否定語	8.000	[1.686, 37.965]	4.089	[1.464, 11.422]	6.933	[1.443, 33.320]
対義語	3.117	[1.174, 8.725]	10.909	[1.337, 89.001]	2.500	[0.719, 8.696]
対比関係	2.549	[0.896, 7.251]	2.753	[0.903, 8.392]	1.529	[0.466, 5.021]
<b>能力低群—能力中群</b>						
実験操作なし	4.235	[1.292, 13.889]	3.360	[1.031, 10.946]	3.106	[0.911, 10.586]
文章中の記述あり	1.517	[0.458, 5.020]	0.574	[0.177, 1.858]	2.113	[0.375, 6.074]
文章中の記述なし	9.412	[2.312, 38.312]	11.636	[2.649, 51.110]	11.566	[2.944, 45.351]
否定語	6.000	[2.086, 17.255]	3.913	[1.421, 10.779]	2.967	[0.909, 9.683]
対義語	1.473	[0.600, 3.617]	3.732	[1.435, 9.704]	1.926	[0.655, 5.680]
対比関係	3.211	[1.207, 8.545]	3.988	[1.500, 10.598]	4.911	[1.628, 14.817]

因果関係の取り違え錯乱枝に関しても、おおむね能力群間の差が顕著に見られたが、文章中の記述なし条件・対比関係条件において能力中群と能力高群の間のオッズの違いが見られず、文章中の記述あり条件で能力低群と能力中群の間のオッズの違いが見られなかった。トレースラインを確認すると、文章中の記述なし条件下での能力中群・能力高群における因果関係の取り違え錯乱枝の選択率はいずれも低かった。これらの結果を踏まえると、英語文章中に書かれていないことがらについて因果関係の取り違えを含んだ錯乱枝は、能力中群以上の受検者を識別する場合には有用でないことが示された。一方、文章中の記述なし条件における低群—中群間のオッズ比は 11.636 倍と非常に大きかったことから、英語文章中に書かれていないことがらについて因果関係の取り違えを含む錯乱枝は、能力低群と能力中群の受検者を識別する場合に有用となる可能性も示唆された。文章中の記述あり条件においては、低群—中群間のオッズ比が 1 に近い値であり、トレースラインを確認すると、能力中群の選択率のほうが能力低群よりも高くなっていた。これらの結果から、文章中の記述あり条件下での因果関係の取り違え錯乱枝は、能力低群・能力中群で同程度の選択率となることが示された。

比較対象の取り違え錯乱枝に関しては、文章中の記述あり条件・文章中の記述なし条件・否定語条件において中群—高群間のオッズ比が大きくなっており、文章中の記述なし条件・対比関係条件において低群—中群間のオッズ比が大きくなっていた。文章中の記述なし条件において、比較対象の取り違え錯乱枝の選択率の差が能力群間で大きく異なることが示された一方で、文章中の記述あり条件・否定語条件においては、能力中群・能力高群のみで選択率が異なっており、能力低群とは異なる選択率であったことが示された。

分析 3-2 を通じて、英語文章中の対比関係を利用した条件に割り当てた項目では、言及対象の取り違えに関する全 3 種類の錯乱枝の選択率が能力低群—中群間で大きくなっていた一方で、能力中群—高群間では大きな違いが見られなかった。また、文章中の記述あり条件を割り当てた項目では、言及対象の取り違えに関する全 3 種類の錯乱枝の選択率が能力低群—中群間では異なっていなかった一方で、能力中群—高群間では選択率が異なるという結果が得られた。

### 2.3.3 考 察

研究 2 では、入学試験問題として出題されていた英語文章読解問題の錯乱枝の特徴として顕著であったものを実際の錯乱枝として反映させ、さまざまな状況下の設問における錯乱枝の選択率を能力群ごとに比較し、錯乱枝の選択率に関する基礎的知見を得ることが目的であった。以下では、それぞれの錯乱枝が有効となる状況・能力群について整理し、問題作成者が多枝選択式問題を作成する際の示唆について述べる。

#### 2.3.3.1 文章中の記述の有無

本研究では、文章中の記述の有無をそのまま錯乱枝に反映させた場合（設問種別 A）と、テスト項目を作成する際の設問の状況として設定した場合（設問種別 B2・B3・C2・C3）に関する分析を行った。分析 1 では、文章中の記述あり錯乱枝・文章中の記述なし錯乱枝ともに、能力中群—高群間では選択率の差が検出されたのに対し、能力低群—中群間では選択率の差が見られなかった。一方で、分析 2-2・分析 3-2 では、能力低群と能力中群との間で文章中の記述なし条件における種々の錯乱枝の選択率が異なるという結果が得られ、能力低群における文章中の記述なし錯乱枝の選択率が能力中群よりも高くなっていたことが示された。

英文法に関する空所補充問題を題材に検討を行った先行研究 (Goodrich, 1977; Morimoto, 2008) では、能力低群において、正答と無関連の語を提示した無関連錯乱枝の選択率が高くなるという知見が得られている。英語文章読解を取り上げた本研究では、無関連錯乱枝に相当するものとして位置づけた文章中の記述なし錯乱枝の選択率の様相が、上記の先行研究とは異なっていた。具体的には上に示したように、文章中に記述されていないことがらを述べるのみの錯乱枝（文章中の記述なし錯乱枝）の選択率は、能力中群と能力高群の間で大きく異なっていたが、文章中に記述されていないことがらを述べた上で、対義語や対比関係による文章中の記述の逆転を含む錯乱枝の選択率は、能力低群と中群の間で異なるという結果が得られた。

上記のような結果が得られた背景には、能力中群の受検者における選択枝の正誤判断のプロセスが、設問の違いに応じて異なっていた可能性が考えられる。英語文章中に記述されていない内容ばかりが錯乱枝として並んでいる設問に解答する場合、能力中群の受検者

は、英語文章中に記述がないという理由の誤答の他に、否定語や対義語、内容の取り違えや因果関係の取り違えなど、他の理由の誤答も含まれた錯乱枝に対して正誤判断を行うことになるため、誤答と判断できる理由が多くなり、結果として文章中の記述なし錯乱枝を誤りと判断できた受検者が多かったことが推察される。一方、文章中の記述あり錯乱枝と文章中の記述なし錯乱枝が同時提示される設問では、文章中の記述なし錯乱枝には「文章中に記述されていない内容である」という誤答のみが反映されているため、能力中群の受検者の多くはその他の誤答の理由を探すことに苦戦を強いられ、結果として文章中の記述なし錯乱枝を誤答と判断するには至らなかったと考えられる。

なお、能力高群では、文章中の記述なし錯乱枝・文章中の記述なし条件の各錯乱枝ともに、選択率が能力中群よりも低かったことから、英語文章中に記述されていないことがらであるという理由のみで選択枝の記述内容を誤答と判断できていた可能性がある。

### 2.3.3.2 対概念の使用

対概念の使用に関して、状況2では実験操作なし条件・文章中の記述あり条件・文章中の記述なし条件の3条件下で、否定語錯乱枝・対義語錯乱枝・対比関係錯乱枝の選択率に関する検討を行った。状況3では、錯乱枝のすべてに否定語が含まれる条件・対義語が含まれる条件・対比関係を利用した表現が含まれる条件を設定し、錯乱枝の選択率に関する検討を行った。以下では、3種類の錯乱枝に関する検討の結果を整理しながら、選択率が高かった理由についての考察を進める。

**否定語錯乱枝** 分析2-2の結果、実験操作なし条件の否定語錯乱枝は能力中群—高群間、能力低群—中群間の両方で選択率の差が検出され、文章中の記述あり条件の否定語錯乱枝は能力中群—高群間で、文章中の記述なし条件は能力低群—中群間でそれぞれ選択率の差が確認された。また、分析2-1の結果、能力中群および能力高群の2群において、文章中の記述あり条件の否定語錯乱枝の選択率が実験操作なし条件よりも高いことが示された。

認知心理学における文処理の研究では、否定語を含む文の処理に時間がかかることが明らかにされている (Carpenter & Just, 1975)。実際、否定語を含まない誤答の文よりも、否定語を含む誤答の文のほうが正誤判断に時間がかかる (Carpenter & Just, 1975)。正誤判断にかかる時間の長さを認知負荷の大きさとして置き換えることが可能であるならば、本研究における否定語錯乱枝は、受検者にとって認知負荷の高い文であったことが考えられる。

文章中の記述なし錯乱枝に対する考察の箇所でも述べたように、文章中の記述なし条件における否定語錯乱枝は、「文章中に記述されていない内容であること」と「文章中に存在する表現に対して否定語が追加されていること」の2つの誤答が含まれているため、能力中群の受検者がこの錯乱枝を誤答であると判断できるポイントが多く、選択率が下がる可能性が考えられる。一方、文章中の記述あり錯乱枝は、英語文章中に記述されていることがらについて否定語が付されているため、能力中群の受検者には内容上問題がないよう



に見え、否定語に気づかずに選択していた可能性が考えられる。また、文章中に記述されている内容について否定語を含む文の処理が難しかったことから、否定語錯乱枝の内容を誤って理解した結果、正答であると判断した可能性も考えられる。

本研究の結果から、能力低群において文章中の記述なし条件の否定語錯乱枝の選択率が高く、能力中群において文章中の記述あり条件の否定語錯乱枝の選択率が高いという結果が得られた。この知見は、問題作成者がテスト項目を作成する際の指針として有効に機能するものと考えられる。

**対義語錯乱枝** 分析 2-2 の結果、文章中の記述なし条件では能力中群—高群間・能力低群—中群間の両方で対義語錯乱枝の選択率の差が確認された。一方、実験操作なし条件では能力中群—高群間のみで選択率の差が、文章中の記述あり条件では、能力低群—中群間のみで選択率の差が検出された。また、分析 2-1 の結果として、能力高群における文章中の記述あり条件の対義語錯乱枝の選択率が、実験操作なし条件に比べて非常に高かったことも明らかになった。能力高群の対義語錯乱枝の選択率を見ても、実験操作なし条件・文章中の記述なし条件では 0% 付近の選択率であったのに対し、文章中の記述あり条件では選択率が高くなっていった。先行研究では、能力低群・中群の受検者に対して対義語錯乱枝が選択されやすいことが報告されているが (Morimoto, 2008)、英語文章読解問題においては設問の状況に応じて異なっていたことが示唆された。

このような結果が得られた背景には、対義語錯乱枝に含まれている対義語の難易度が関連しているものと考えられる。能力高群の受検者の場合、英語文章中に long (長い) と記述されている状況下で、錯乱枝中に short (短い) という記述があれば、誤りに気づくことは難しくないと想像される。一方で、本研究の対義語錯乱枝は、英語文章中に mad (狂気) と記述されている状況下で、対義語錯乱枝中に sane (正気) を提示した設問が存在した。易しいレベルの対義語の場合はすぐに誤答と判断できるが、難易度の高い対義語が提示されていた場合には誤答と判断できないケースもあることが推察される。実際、英語文章中の文に含まれる語自体の難易度が高いため、その語の対義語はさらに難易度が高かったり、馴染みのない語であったりすると想定される。このことを考慮すると、文章中の記述あり条件で対義語錯乱枝を作成する場合には、語の選定に十分注意する必要があると考えられる。

**対比関係錯乱枝** 分析 2-2 の結果、文章中の記述あり条件・文章中の記述なし条件の対比関係錯乱枝について、能力低群では能力中群の 7 倍程度選択されやすいことが示された。また、分析 2-1 の結果においても、能力低群で文章中の記述あり条件・文章中の記述なし条件ともに、実験操作なし条件よりも対比関係錯乱枝の選択率が高いことが示された。これらの結果から、能力低群においては、文章中に記述されている内容であるかどうかを問わず、対比関係錯乱枝の選択率が高くなることが示された。また、能力中群—高群間では、実験操作なし条件・文章中の記述なし条件で選択率の差が検出された。さらに、

分析 3-2 においても、対比関係条件下の内容の取り違い錯乱枝・因果関係の取り違い錯乱枝・比較対象の取り違い錯乱枝のすべてにおいて、能力低群—中群間の選択率の差が検出された。概して、文章中の対比関係を利用した錯乱枝は、能力低群の受検者において選択率の高い錯乱枝であったことが明らかとなった。

上のような結果が得られた背景には、英語文章中で「何と何を比較しているか」について、能力低群の受検者が十分理解できていない可能性が考えられる。例えば、冊子 4 ではルネサンス期の芸術家について記述した英語文章を提示し、対比関係錯乱枝として“*Michelangelo was an exception of Renaissance artists*”（ミケランジェロはルネサンス期の芸術家の中では例外である）を提示した。英語文章中では、ルネサンス期の芸術家が自己省察のために孤独・秘密を好んだという記述があり、ミケランジェロもそのような生活をしてきたことが述べられていたことから、上の選択枝は誤答となる。このような設問において対比関係錯乱枝の正誤を判断する際、ルネサンス以前の芸術家とルネサンス期の芸術家の特徴を英語文章中に記述されている範囲でそれぞれ整理し、ミケランジェロがどちらの特徴をもっているか判断することになると考えられる。この過程の中で、能力低群の受検者はルネサンス以前の芸術家の特徴とルネサンス期の芸術家の特徴の整理が十分でなかったために、ミケランジェロがルネサンス期の芸術家の中では例外であるとする対比関係錯乱枝を選択したものと考えられる。

一方で、能力中群・高群においては、文章中の記述あり条件における対比関係錯乱枝の選択率が低かったことから、中程度以上の能力水準にある受検者は対比されていることからの整理が十分にできたものと推察される。

### 2.3.3.3 言及対象の取り違い

言及対象の取り違いに関して、実験操作なし条件・文章中の記述あり条件・文章中の記述なし条件・否定語条件・対義語条件・対比関係条件の 6 条件下で、内容の取り違い錯乱枝・因果関係の取り違い錯乱枝・比較対象の取り違い錯乱枝の選択率に関する検討を行った。以下では、3 種類の錯乱枝に関する検討の結果を整理しながら、選択率が高かった理由についての考察を進める。

**内容の取り違い錯乱枝** 分析 3-1 の結果、設問種別間で内容の取り違い錯乱枝の選択率の違いは特に見られないという結果が得られた。一方、分析 3-2 の結果、大部分の条件において能力中群—高群間、能力低群—中群間の選択率の差が検出された。例外として、文章中の記述あり条件・対義語条件では能力低群—中群間で選択率が大きく異ならず、対比関係条件では能力中群—高群間で選択率が大きく異ならないという結果が得られた。能力低群と中群との間で選択率の違いが見られなかった上記 2 つの条件では選択率の様相がそれぞれ異なり、文章中の記述あり条件では能力低群・中群ともに内容の取り違い錯乱枝の選択率は低かったのに対し、対義語条件では両群ともに選択率が一定程度高かった。

上記の結果から、内容の取り違い錯乱枝は一部の条件を除き、能力群間で選択率の違いが顕著となる錯乱枝であることが示唆された。受検者の能力水準が高くなった場合に、英語文章中の内容を取り違えた記述に対して誤答と判断することができるものと考えられる。

**因果関係の取り違い錯乱枝** 分析 3-1 の結果、能力中群での因果関係の取り違い錯乱枝に関して、実験操作なし条件に比べて文章中の記述あり条件の選択率のほうが高いことが示された。また、分析 3-2 の結果から、この条件における因果関係の取り違い錯乱枝については、能力中群—高群間の選択率の差も確認された。これらの結果を踏まえると、文章中の記述あり条件における因果関係の取り違い錯乱枝は、能力中群内で選択率が高く、かつ能力高群と比べた場合も選択率が高いということが明らかにされた。加えて、能力低群—中群間では大きなオッズ比が得られなかった点を考慮すると、文章中に記述された内容について因果関係を取り違えた錯乱枝は、能力低群・中群と高群の受検者の能力を識別するものであったことが示された。一方、文章中の記述なし条件における因果関係の取り違い錯乱枝に関しては、文章中の記述あり条件とは異なり、能力低群—中群間の選択率の差が検出された。以上の結果をまとめると、文章中の記述なし条件における因果関係の取り違い錯乱枝は能力低群で選択率が高く、文章中の記述あり条件における因果関係の取り違い錯乱枝は能力低群・中群で選択率が高いことが明らかにされた。

錯乱枝の理論的分類に関する先行研究 (King et al., 2004; Lin et al., 2010) では、英語文章読解テストにおける選択枝を 4 つに分類しているが、上記 2 条件における錯乱枝を理論的分類と対応づけて考えると、文章中の記述なし条件における因果関係の取り違い錯乱枝は Level 1「文章に関連しない内容を含む誤答選択枝」に相当し、文章中の記述あり条件における因果関係の取り違い錯乱枝は Level 2「文章中に記述されている事実どうしの関係を取り違えている誤答選択枝」に相当するものと考えられる。このような位置づけの中で本研究の結果を解釈すると、能力低群の受検者は、因果関係の判断を行う以前に文章中の記述の有無を判断することが難しかった一方で、能力中群の受検者は、選択枝中の内容に関連した記述が英語文章中に存在することは理解できていたが、どちらが原因（先行する事象）でどちらが結果（後続する事象）であるかを十分に判断できなかったため、二つのことからの時間的な経過や論理関係の判断を誤った可能性が考えられる。

**比較対象の取り違い錯乱枝** 分析 3-1 の結果、能力高群のみにおいて、文章中の記述なし条件・否定語条件における比較対象の取り違い錯乱枝の選択率が、実験操作なし条件の選択率に比べて低くなることが示された。また、分析 3-2 において、文章中の記述なし条件・否定語条件下の錯乱枝の選択率は能力中群—高群間で大きな差となっていたことから、能力高群において極めて選択率の低い錯乱枝であったことが明らかとなった。一方、能力低群と能力中群との間の選択率の差に関しては、文章中の記述なし条件・対比関係条件で顕著であった。

文章中の記述なし条件は、本研究におけるその他の多くの錯乱枝に関しても能力群間の選択率の差が検出されており、英語文章中に記述されていないことに対する正誤判断そのものが、能力群間の差をもたらしたものと推察される。また、否定語条件は、内容の取り違い錯乱枝・因果関係の取り違い錯乱枝においても能力中群—高群間の選択率の差が確認されていたことから、能力高群の受検者の多くは否定語を含む錯乱枝に対して誤りであると判断できることが示唆された。対比関係条件における比較対象の取り違い錯乱枝については、対比関係錯乱枝の箇所述べたように、能力低群の受検者において文章中の対比構造を整理できていなかったために、能力低群のみで高い選択率となっていたことが考えられる。

## 2.4 第2章のまとめ

本章では、英語文章読解問題における錯乱枝の基礎的研究として位置づけた研究1・研究2を実施し、1) 錯乱枝に反映されている誤答の特徴の把握、2) 能力水準ごとにみた錯乱枝の選択率の様相の2つの課題の解決に取り組んだ。以下では、上記2つの研究課題に即して、研究1・研究2から得られた知見を整理する。

第1に、錯乱枝に反映されている誤答の特徴の把握に関しては、研究1において大学入学試験問題の英語文章読解問題を材料に調査を行った結果、計8種類の誤答の特徴を明らかにした。加えて、計8種類の誤答の特徴は「文章中の記述の有無」・「対概念の使用」・「言及対象の取り違い」の3つに大別できることも示された。具体的には、文章中に記述されている内容に関する誤答や、文章中には記述されていない内容が選択枝中に含まれているという誤答が「文章中の記述の有無」というカテゴリに、否定語や対義語を含んでいるという誤答、文章中に述べられている対比を利用した誤答が「対概念の使用」というカテゴリに、文章中の内容や比較対象・因果関係を取り違えた誤答が「言及対象の取り違い」というカテゴリに分類できた。それぞれの誤答の特徴の多くは、先行研究における研究知見や理論的分類と対応するものであり、誤答選択枝に反映された誤答の特徴として一定程度妥当であると考えられる。

第2に、それぞれの錯乱枝の選択率が高くなる能力水準の検討に関しては、研究2において以下の3点が明らかとなった。1つ目に、能力低群の受検者には、文章中に記述されていないことについて否定語を含めたり、文章中の対比を利用した表現と組み合わせたり、因果関係を取り違えたりした錯乱枝を作成した場合に選択率が高くなる傾向が確認された。2つ目に、能力中群の受検者には、文章中に記述されている内容に否定語を含めたり、因果関係を取り違えたりした錯乱枝を作成した場合に選択率が高くなる傾向が見られた。3つ目に、能力高群の受検者には、文章中に記述されている内容に対義語を加えた錯乱枝を作成した場合に選択率が高くなることが示された。

以上2点の課題に取り組むことにより、選択枝の正誤判断の際の受検者の特徴を考慮した錯乱枝の作成に資する知見が得られ、問題作成者がテスト項目を作成するときの参考資料を提供した点において、本章における研究1・2は意義のあるものであると考えられる。加えて、本章において選択率の高さが示された錯乱枝のうち、受検者が解答しているときの認知過程を推測できるような、独特な機能をもつものも複数検出された。

本章における2つの研究は、英語文章中に記述されていない内容を含めた錯乱枝が過去の入学試験問題において多く用いられていたことを明らかにしたと同時に、そのような錯乱枝の選択率が能力群間で大きく異なることも実証的に示し、これまでによく用いられてきた錯乱枝の機能についてエビデンスを追加する役割を果たしたものと考えられる。加えて、英語文章中の記述に否定語や対義語を追加した錯乱枝は、過去の入学試験問題ではあまり用いられていなかったが、設問の状況によっては有効に機能する可能性があることも示した。過去の入学試験問題で取り上げられることの少なかった誤答についても、受検者の誤答の特徴を具体的に示し、錯乱枝の機能に関する知見を提供したものと考えられる。



## 第3章

# 英語文章読解テスト解答時の情報処理プロセスに基づく錯乱枝開発

### 3.1 研究の背景

第2章で行った研究1・2では、Embretson & Wetzel (1987) のモデルにおける「選択枝の正誤の評価」に焦点を当て、錯乱枝の種類の違いと受検者の正誤判断の様相について検討を行い、問題作成者がどのような錯乱枝を作成すればよいかということに対する指針を提供しうる知見を得た。ただし、第2章では Embretson & Wetzel (1987) が提唱したモデルの中で「選択枝の正誤の評価」の部分のみを扱ったにすぎず、受検者の認知過程のごく限られた部分のみを錯乱枝に反映させた場合の知見の提供にとどまっている。多枝選択式テストの構造上、各選択枝は設問の下に置かれるため、その設問においてどのような読解を求めるかということによって、錯乱枝への正誤判断が異なり、その結果、選択率が異なる可能性も考えられる。第3章では、Embretson & Wetzel (1987) のモデルの「文章のマッピング」と「選択枝の正誤の評価」の双方を考慮に入れた検討を行うことで、受検者が英語文章読解問題に解答する際に生じる誤りをより詳細に反映させた錯乱枝の作成手法を提案できる知見の提供を目指す<sup>2</sup>。

設問と選択枝の関係を検討するにあたっては、「受検者が英語文章読解問題をどのように解くか」という問いから出発することが必要である。第1章でも述べた通りであるが、文章読解問題を解く際の受検者の情報処理過程に関するモデル (Embretson & Wetzel, 1987; Embretson, 2001) では、それぞれの選択枝の内容に関連する記述を文章中で特定する「文章のマッピング」という過程を経て、選択枝の正誤判断を行うものとしている。受検者は設問を読んだ後、それぞれの選択枝の内容に関わりの深い記述を文章中で特定した上で、選択枝の正誤判断を行うことが仮定される。先行研究では、その文章読解項目に正答するために読解する必要のある文をキーセンテンス (key sentences) と呼んでおり、正答選択枝

<sup>2</sup> 本章の一部は、寺尾他 (2017) を加筆・再構成したものである。

の内容に関連する文，およびそれに隣接する 2~3 文を指すものとしている (Carr, 2006). 本章では，受検者が項目に正答するにあたって，正答選択枝だけではなく，誤答選択枝の内容に関連する文やそれに隣接する文の読解も必要となる点に着目し，誤答選択枝も含めた選択枝それぞれの内容に関連する文，あるいはそれに隣接した 2~3 文をキーセンテンスと呼ぶこととした．なお，キーセンテンスという用語は，正答のために必要な情報 (necessary information) という呼び方がなされる場合もあるが (e.g., Buck et al., 1997), 両者はほぼ同一の定義をもつ用語であるため，以下では「キーセンテンス」と呼ぶ先行研究と「正答のために必要な情報」と呼ぶ先行研究の区別を行わずにレビューした．

本章では，受検者が設問・選択枝を読んでから選択枝の正誤判断を行うまでの間に生じる認知過程のうち，次の 2 点に着目した．第 1 に，各選択枝のキーセンテンスを特定する過程，第 2 に，キーセンテンスと選択枝を見比べて両者の異同を吟味する過程である．

1 つ目に挙げた「各選択枝のキーセンテンスを特定する過程」では，それぞれの選択枝の内容を理解した上で，内容的に関連するキーセンテンスを見つけることが必要となる．受検者がそれぞれの選択枝の内容に対応するキーセンテンスを探す場合，複数存在するキーセンテンスが文章中のまとまった箇所が存在するような設問なのか，文章全体に散らばるような設問なのかということは，選択枝の正誤判断に大きな影響を及ぼしうる．Embretson & Wetzel (1987) では，まとまった箇所に存在するキーセンテンスは比較的特定しやすく，文章全体に散らばっているキーセンテンスの特定は難しいということが指摘されている．また，Buck et al. (1997) は，TOEIC の読解問題に対する解答の背後に文章全体に散らばっているキーセンテンスを統合して理解するという下位スキルが見い出されたことを報告している．

受検者に対して，どのような範囲にあるキーセンテンスの特定を求めるかということ (キーセンテンスの位置関係) は，設問の設定の仕方に依拠するものと考えられる．一杉 (2007) は，キーセンテンスという用語こそ用いていないものの，キーセンテンスの位置関係の観点から，大学入試で出題される英語文章読解項目を「部分一致問題」「段落一致問題」「全体一致問題」の 3 つに分類している．「部分一致問題」は英語文章中の一部分を踏まえて解答する問題，「段落一致問題」は英語文章中の特定の 1 段落全体の記述を踏まえて解答する問題，「全体一致問題」は英語文章全体の記述を踏まえて解答する問題である (一杉, 2007). 本章では一杉 (2007) の分類を参考に，キーセンテンスの位置関係が異なる設問を複数種類作成し，それらの設問の下で錯乱枝の選択率に関する検討を行う．

また，2 つ目として挙げた「キーセンテンスと選択枝を見比べて両者の異同を吟味する過程」では，キーセンテンスと選択枝に共通して含まれる特徴が，選択枝の正誤判断に影響を及ぼすものと考えられる．先行研究では，キーセンテンス中の語と選択枝中で使用される語が選択率に及ぼす影響について検討されているが，得られた結果が一致していないという状況にある (e.g., Buck et al., 1997; Drum et al., 1981; Ushiro et al., 2007). 具体的には，



キーセンテンス中で用いられている語を錯乱枝中で用いたとき、その錯乱枝が選ばれやすくなるという結果を報告している先行研究がある一方で (Drum et al., 1981; Ushiro et al., 2007), キーセンテンス中の語と錯乱枝中の語の重複があると、その錯乱枝が選ばれにくくなるという結果を報告する先行研究 (Buck et al., 1997) も存在する。

このような研究結果の不一致は、英語文章読解問題における錯乱枝の効果的な作成を妨げる可能性がある。この不一致を解決する要因の一つとして、設問タイプが挙げられる。Freedle & Kostin (1993) は、設問タイプに応じて語の重複が正答選択枝の選択率を高めるかどうか異なることを報告している。実際、文章中に明示されている情報について問う設問では、すべての受検者において語の重複が正答選択枝の選択率を高めるのに対し、文章全体の構造やテーマの理解・文章から導き出される推論について問う設問では、能力の低い受検者に対してのみ語の重複が正答選択枝の選択率を高め、中程度から高い能力水準にある受検者に対して語の重複の効果は検出されなかった (Freedle & Kostin, 1993)。なお、のちに清水 (2005) が設問タイプの整理を行っており、前者の設問を下位レベル設問 (lower-level question), 後者の設問を上位レベル設問 (upper-level question) と呼んでいる。

このように、キーセンテンスと選択枝での語の重複が正答選択枝の選択率に及ぼす影響は、設問タイプに応じて異なることが示されている。ただし、先行研究において、錯乱枝の選択率を取り上げたものは見られない。そこで本章では、キーセンテンスと錯乱枝との間での語の重複が錯乱枝の選択率に及ぼす影響について、設問タイプの違いを考慮し、能力群別に検討を行うこととした。

本章では、上記に挙げた2つの認知過程に対応させて、以下の2つの研究課題を設定した。第1の研究課題は、各選択枝の内容に関連するキーセンテンスを見つけ出す過程に着目し、キーセンテンスの位置関係が異なる設問を用いてキーセンテンスの特定の容易さを間接的に操作し、各設問の錯乱枝の選択率について、能力群別の検討を行うことである。第2の研究課題は、キーセンテンスと選択枝を見比べて両者の異同を吟味する過程に着目し、キーセンテンスと選択枝の語の重複の有無が異なる設問を用いてキーセンテンスと選択枝の比較の容易さを間接的に操作し、錯乱枝の選択率について設問タイプ別・能力群別の検討を行う。

第1の研究課題を解決するにあたって、研究3・研究4の2つの研究を行った。3.2節では研究3として、キーセンテンスの特徴を把握するための調査を行い、英語文章読解問題におけるキーセンテンスの位置関係について整理する。3.3節では研究4として、研究3において整理した結果を踏まえ、キーセンテンスの位置関係を操作した設問における錯乱枝の選択率について、能力群別の検討を行った。第2の研究課題を解決するため、3.4節では研究5として、キーセンテンス中に含まれる語を錯乱枝にも用いる条件と、難易度の低い類義語に置き換えた条件を設定し、下位レベル設問と上位レベル設問における錯乱枝の選択率を能力群別に検討した。3.5節では本章のまとめとして、本章で得られた知見およ

び示唆を整理する。

なお、本章では、第2章で独特の効果が確認された「否定語錯乱枝」・「対義語錯乱枝」・「因果関係の取り違い錯乱枝」の3種類を取り上げることとした。第2章では、否定語錯乱枝・因果関係の取り違い錯乱枝に関して、文章中の記述あり条件で能力中群の選択率のほうが能力高群よりも高いこと、対義語錯乱枝に関して、文章中の記述あり条件で能力低群の選択率のほうが能力中群よりも高く、能力高群でも一定程度の選択率があることが明らかとなっている。これらの知見から、否定語錯乱枝・因果関係の取り違い錯乱枝は能力低群・中群における誤答をターゲットとしており、対義語錯乱枝は能力低群における誤答をターゲットとしていることが示唆される。ここで、第2章の知見を文章中の記述あり条件に限定して述べたのは、本章がキーセンテンスの特定や、キーセンテンスと選択枝との比較検討を念頭に入れているためである。キーセンテンスに否定語や対義語を追加したり、キーセンテンス中の因果関係を取り違えて記述したりした錯乱枝について、キーセンテンスの特定や選択枝との比較検討を操作した上記の設問設定のうち、どのような状況下でより選択率が高まったり、能力群間の選択率に大きな差が生じるのかといった点を検討することは、項目作成の現場を支援するエビデンスを提供するものと考えられる。

## 3.2 研究3：キーセンテンスの特徴に関する検討

研究3では、キーセンテンスの位置関係に関するカテゴリを整理するため、大学の入学試験問題として使用された英語文章読解問題を題材に、キーセンテンスの位置関係に関する調査を行う。

### 3.2.1 方法

#### 3.2.1.1 研究協力者

心理学を専攻する大学院生4名（男性1名、女性3名）に研究協力を求めた。研究協力者4名は、計量心理学および教育測定学の講義を受講した経験があり、テストの作成に関して一定程度の知識を有するものであった。

#### 3.2.1.2 材料

大学の入学試験問題のうち、多枝選択式の英語文章読解問題を計17題使用した。英語文章読解問題の選定にあたっては、以下の3点に留意した。第1に、英語文章の内容そのものの理解を求める問題であり、かつキーセンテンスが文章中に存在すると考えられる問題を選定した。入学試験の英語文章読解問題の中には、英語文章の内容理解を求める問題のほかに、英語文章中の単語の意味について問う問題や、英語文章に対する適切なタイトルを選ぶよう求める問題などもある。このような問題を用いた場合、キーセンテンスの位

Table 3.1  
研究3 で用いた文章のリーダビリティ

文章 No.	文章の内容	リーダビリティ	設問数
冊子 1-1	コモン・センスと公正	8.21	10
冊子 1-2	移民と食事	14.03	5
冊子 1-3	長生きの秘訣	8.86	4
冊子 2-1	問題の発見と解決	9.73	3
冊子 2-2	インドとアメリカの経済の違い	10.33	4
冊子 2-3	世界経済	8.02	5
冊子 2-4	金剛組の破産	15.14	2
冊子 2-5	言語権	13.57	3
冊子 2-6	無声映画とベストアクター賞	13.47	4
冊子 3-1	フリースクール	13.19	4
冊子 3-2	スポーツの標準化	12.42	4
冊子 3-3	結婚の多様化	15.89	2
冊子 3-4	ジャンヌダルク	9.25	3
冊子 3-5	メディアと報道	13.51	4
冊子 3-6	知能と情動・共感	11.98	4
冊子 4-1	身体感覚と信念	8.75	10
冊子 4-2	卵のしくみ	9.44	6

置関係に関する検討を行うことが困難になる可能性があることを考慮して、本研究の検討対象から除外した。第2に、本研究の調査で取り扱う問題全体が後述する分類カテゴリを網羅できるよう、英語文章を多数使用した。キーセンテンスの位置関係は、英語文章や設定される設問に応じて異なることが想定されるため、少数の英語文章読解問題を用いた場合には、特定のカテゴリにあてはまる問題が存在しない可能性がある。第3に、検討対象の英語文章の難易度にバリエーションをもたせるため、英語文章のリーダビリティを算出し、さまざまな難易度の文章を選定した。リーダビリティの算出は第2章と同様、Flesch-Kincaidの公式を用いた。研究3で使用した文章の一覧をTable 3.1に示した。

### 3.2.1.3 キーセンテンスの特徴を把握するための分類カテゴリ

一杉(2007)の分類を参考にして、英語文章中のキーセンテンスの位置関係に関する分類カテゴリを作成した。一杉(2007)の分類に対応させる形で、「部分一致問題」を「一文を踏まえる」、「段落一致問題」を「一段落を踏まえる」、「全体一致問題」を「文章全体を踏まえる」とした。さらに、選択枝の正誤判断に関連する情報が英語文章中の複数の段落にまたがる場合があることも想定し、上の3つのカテゴリに加えて「複数段落を踏まえる」を設定した。

#### 3.2.1.4 調査冊子の構成

使用した英語文章読解問題を研究協力者4名に割り当てることができるよう、計4種類の調査冊子を作成した。調査冊子1種類あたりの設問数が20問前後になるよう、英語文章読解問題の調査冊子への割り当てを調整した。調査冊子では、キーセンテンス・設問・正答選択枝のみを提示し、その下にキーセンテンスの位置関係に関する評価欄を設けた。本研究で使用した調査冊子の例を Figure 3.1 に示した。

研究協力者の認知的負荷を必要最小限とするため、以下の3点に留意して調査冊子を構成した。第1に、調査冊子に英語文章全体を掲載することはせず、英語文章中からキーセンテンスのみを抽出して提示した。調査対象者には、1名あたり設問20問程度の評価を求めるため、英語文章全体を提示し、それを通読した上で評価を求めることは、調査協力者にとって多大な負担であると考えられる。加えて、英語文章全体を提示した場合、その項目への解答にあたって必要のない情報を提示することになり、キーセンテンスへの評価に影響を及ぼす可能性がある。過不足なくキーセンテンスのみを提示することで、研究協力者1名あたりの負担を最小限にし、キーセンテンスの評価を効率的に行ってもらえることが可能となる。このような対応の影響で、調査協力者が英語文章全体の中での当該のキーセンテンスの位置を把握できなくなることに配慮し、それぞれのキーセンテンスが何段落目の何文目であるかを、「第○段落・△文目」といった形で付記することとした。

第2に、選択枝に関してもキーセンテンスと同様、研究協力者の負担を最小限にするため、正答選択枝のみを提示し、誤答選択枝は提示しなかった。このような対応の背景として、誤答選択枝の中には対応するキーセンテンスをもたないものもあったことが挙げられる。本研究では、すべての設問において正答選択枝のみを提示することとした。

第3に、研究協力者の認知的負荷を軽減する目的から、英語で書かれたキーセンテンス・設問・選択枝に1対1対応させる形で、研究協力者の母語である日本語の訳を付した。その理由としては、本研究がキーセンテンスの特徴の把握を目的としており、研究協力者の英語文章読解能力の測定が目的ではなかったことが挙げられる。

#### 3.2.1.5 手続き

研究協力者による評価は、2014年7月～8月の期間で実施された。本研究の調査目的を伝えた上で評価方法について説明し、評価結果に対する確認と聞き取りを行うため、各研究協力者へのインタビューを計2回行った。1回目のインタビューでは、英語文章読解問題に関する調査であることを研究協力者に伝え、研究の目的および調査の概要を説明した上で、研究参加への同意を確認した。その後、調査冊子の構成や各カテゴリの定義について説明を行い、調査冊子中にある評価例を参照しながら、評価方法の確認を行った。研究協力者が評価するにあたり、原則として上述の4つのカテゴリからあてはまるものを選択

## **Section 16**

本文(全8段落)

第2段落・4文目

One important aspect of optimism is the way optimists think about the causes of events.

楽観主義の重要な側面の一つは、出来事の原因に関する楽観主義者の考え方である。

第2段落・5文目

They tend to believe that a good event is the result of their own skill or effort, but that a bad event is the result of chance.

彼らは、よい出来事は自分の技術や努力の結果であると考え、悪い出来事は偶然の結果であると考える傾向がある。

第2段落・6文目

This gives them hope to keep on trying.

このことが、彼らに努力し続ける希望を与える。

問 According to the second paragraph, starting "The first healthy," optimists

第2段落によると、楽観主義者は … .

正答 usually think they deserve the good things that happen to them.

たいてい、彼らはいいことが起こる価値のある人間だと考える。

解答を導くために求められる文章の読み

- ( ) 一文を踏まえる
- ( ) 同一段落中の複数の箇所を踏まえる
- ( ) 複数の段落の複数の箇所を踏まえる
- ( ) 文章全体を踏まえる

Figure 3.1. 研究3で使用した調査冊子の例

するよう求めた上で、どのカテゴリにもあてはまらないものがあつた場合や、カテゴリの定義に違和感があつた場合には、調査冊子の余白に書き留めておくよう求めた。1回目のインタビューから2週間前後経過した日に2回目のインタビューを設定し、それまでの間に調査冊子に収録されているすべての設問への評定を終えるよう求めた。なお、著者も研究協力者と同じタイミングで調査冊子4種類すべてに対して独立に評定を行い、2回目のインタビューにおいて両者の評定を突き合わせられるようにした。2回目のインタビューでは、各研究協力者と著者で評定結果を設問ごとに確認した。確認の際、新しい分類カテゴリや、既存の分類カテゴリの定義の変更が必要な箇所があれば提案するよう求めた。2回目のインタビューは、研究協力者1名あたり1時間～1時間半程度行われた。

本研究を含めた第3章の研究は、名古屋大学大学院教育発達科学研究科の研究倫理審査委員会の承認を受けて実施された (ERB No. 14-537)。

### 3.2.2 結果と考察

各研究協力者に対する2回目のインタビューで得られた意見・コメント等をもとに、カテゴリの修正を行った。研究協力者との2回目のインタビューでは、キーセンテンスの位置関係についておおむね迷いなく評定できたとの回答が得られた。研究協力者4名のうち、「一文を踏まえる」のカテゴリに関する意見が1件あり、「提示されているキーセンテンス1つを読めば正答選択枝を選ぶことは可能であるが、関連するキーセンテンスをもう1つ確認することで、確実に正答選択枝を選ぶことができる」という指摘があつた。本研究では、一杉(2007)の分類に対応させる形で「一文を踏まえる」というカテゴリを設定したが、第二言語習得の領域における先行研究では、キーセンテンスを1文のみに限定しておらず、隣接する文も含めて定義している(Carr, 2006)。以上のことを踏まえ、「一文を踏まえる」カテゴリを「一内容を踏まえる」カテゴリに変更し、隣接する2～3文を含めた箇所にキーセンテンスが集約されている状況を指すものとした。その他のカテゴリに対しては、特に修正を加えなかつた。

## 3.3 研究4：キーセンテンスの特徴と錯乱枝の選択率との関連

研究4では、研究3で整理されたキーセンテンスの位置関係に関するカテゴリに基づき、キーセンテンスの位置関係が異なる設問における錯乱枝の選択率について検討を行う。

### 3.3.1 方法

#### 3.3.1.1 受検者

愛知県・東京都の大学に通う大学生427名(男性130名、女性297名)を受検者とした。

### 3.3.1.2 材 料

これまでと同様、大学の入学試験問題のうち、多枝選択式の英語文章読解問題を 16 題使用した。研究 4 では、目的にあわせて設問や選択枝を作成するため、英語文章のみを使用した。使用した文章の一覧を Table 3.2 に示した。リーダビリティの指標は、これまでと同様 Flesch-Kincaid の公式により算出した。使用した英語文章のリーダビリティの平均は 10.92、標準偏差は 2.37 であった。

Table 3.2  
研究 4 で用いた文章一覧

文章 No.	文章の内容	リーダビリティ	語数
共通項目	人間には変化が必要	7.91	257
冊子 1	教育が予測するもの	12.07	456
冊子 2	物理学の発展	10.96	495
冊子 3	アフリカの貧困とその報道	12.77	508
冊子 4	メタ・メッセージと他者への配慮	9.07	556
冊子 5	シカゴの南側とホワイト・フライト	11.24	544
冊子 6	言葉の伝達	10.62	472
冊子 7	エコツーリズム	8.99	510
冊子 8	友達	8.00	440
冊子 9	考古学と DNA 解析	15.61	433
冊子 10	批判	14.94	450
冊子 11	リテラシー	13.27	398
冊子 12	ゲームで Java を学習する	10.99	443
冊子 13	睡眠	8.21	547
冊子 14	ペルム紀における生物の絶滅	8.96	485
冊子 15	研究	11.17	455

### 3.3.1.3 設問の作成

本研究では、研究 3 で整理されたキーセンテンスの位置関係に関するカテゴリに沿って設問を作成した。Table 3.2 に示した 16 題の英語文章のうち、共通項目として使用するもの以外の 15 題の英語文章を使用し、英語文章 1 題につき 4 項目の実験項目、計 60 項目を作成した。各設問では、正答選択枝・否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の計 4 枝を作成した。これらの誤答選択枝は、研究 2 において独特の効果が示されたものである。

それぞれの英語文章に付随する読解問題 4 項目の内訳は、キーセンテンスの位置関係に関するカテゴリに対応させ、「一内容を踏まえる」設問、「同一段落を踏まえる」設問、「複数の段落を踏まえる」設問、「文章全体を踏まえる」設問であった。「一内容を踏まえる」

問2 Printing technology 

5
---

.

- 正答 ① was first invented around A.D.100, but making copies was boring and took time  
因果 ② enabled us to make massive copies in a short time after Morse's telegraph was invented  
否定 ③ was outstanding, but multiple and cheap copies were not common to ordinary people  
対義 ④ became popular in the modern era, but people who wanted to learn how to read letters were rare

- (2) Writing enabled people to fix words on a surface, but it did not provide the basis for making multiple, inexpensive copies of materials. ① Additional copies of writings required long and tedious work by scribes, people who copied documents by hand. Consequently, the writings of earlier times were not available to most people. ② Printing made it possible to produce thousands of copies in less times than it had taken a scribe to produce one. ③ Inexpensive written materials became widespread among ordinary people. ① The origin of printing dates back as far as A.D. 100, but the modern era of printing began when Johannes Gutenberg created movable letters from which words could be formed. ④ The single invention of printing encouraged more and more people to learn to read as well as to read to learn.
- (3) ② Although writing and printing represented a major breakthrough in the way that messages were converted into codes, they did not do so in the way that messages were transmitted. In the 1700s a written message still had to be transported by conventional means such as horse cart. But in the early 1800s inventors made great progress in sending symbols via electrical impluses over wires. ② By 1832 Samuel F. B. Morse had invented the telegraph and Morse's telegraph spanned the 60-kilometer distance between Washington and Baltimore by 1844. The telegraph had, at long last, freed long-distance communication from transportation. Electronic advances in the 1990s made it possible to transmit messages without the use of wires. In broadcasting, messages are converted into signals on electromagnetic waves that travel through space. By 1901 coded messages were sent across the Atlantic Ocean by wireless telegraph (early radio) . Messages could be

Figure 3.2. 研究4における設問と対応するキーセンテンスの例



設問では、各選択枝の正誤判断に必要なキーセンテンスが同一段落中の一文、あるいは隣接した2~3文になるように選定し、キーセンテンスに基づいて各選択枝を作成した。この設問では、異なる選択枝に対して異なるキーセンテンスが対応するように作成することを原則としたが、原則を守ることが難しい場合には、キーセンテンスが選択枝間で重複することも許容することとした。「同一段落を踏まえる」設問は各選択枝のキーセンテンスが同一段落中の文となるように作成し、「複数段落を踏まえる」設問は各選択枝のキーセンテンスが2つ以上の段落にまたがるように作成した。「文章全体を踏まえる」設問では、「複数段落を踏まえる」設問以上の広い範囲にキーセンテンスが存在するように作成した。Figure 3.2に、複数段落を踏まえる条件の設問例と対応するキーセンテンスの例を示した。

#### 3.3.1.4 テスト冊子の構成

本研究で独自に作成した英語文章読解問題15題に対応させる形で、計15種類のテスト冊子を作成した。それぞれのテスト冊子は、第1問・第2問の大問2問構成であり、いずれも多枝選択式項目とした。第1問では3項目、第2問では4項目を提示した。研究2と同様、本研究でも共通項目をアンカーとした等化を念頭に入れ、各テスト冊子の大問2問のうち、第1問はすべての種類のテスト冊子を通じて同一の英語文章・多枝選択式問題を提示する共通項目としたが、第2問では、15題の英語文章のうち1題を1種類のテスト冊子に割り当てた。各テスト冊子における第2問の多枝選択式問題4項目には、「一内容を踏まえる」設問・「同一段落を踏まえる」設問・「複数の段落を踏まえる」設問・「文章全体を踏まえる」設問がそれぞれ1項目ずつ含まれるようにした。このようなデザインとすることにより、受検者全員から全4種類の設問に対する解答を得ることができ、「キーセンテンスの位置関係」に関する要因を被験者内（参加者内）要因として取り扱うことが可能となるようにした。

#### 3.3.1.5 手続き

解答データの収集は、2014年10月～2015年1月までの期間で行われた。受検者には、このテストが受検者自身の能力を測定しようとする目的で行われるものではなく、テスト項目の性質について検討する目的で行われるものであることを伝えた。また、受検者に対する倫理的配慮として、テスト冊子が研究室内で厳重に保管されること、研究参加を辞退しても受検者自身が不利益を被ることがないことなどを確認した後、受検者の同意を得た。テストに解答する際、正しいと思う選択枝の丸番号を塗りつぶす形式で解答するよう受検者に教示した。解答時間は30分間であった。

Table 3.3  
項目分析の結果

第1問		項目	正答率	点双列相関					
共通項目 (N = 427)	問1		.585	.477					
	問2		.319	.331					
	問3		.515	.273					
第2問		項目	選択枝	選択率	点双列相関	項目	選択枝	選択率	点双列相関
冊子1 (27名)	問1	1 (否定)		.148	-.271	冊子3 (29名)	1 (対義)	.107	-.390
		2 (正答)		.630	.428		2 (因果)	.179	-.376
		3 (対義)		.111	-.229		3 (正答)	.679	.448
		4 (因果)		.111	-.389		4 (否定)	.036	-.174
	問2	1 (対義)		.296	-.195		1 (否定)	.143	-.263
		2 (否定)		.222	-.349		2 (対義)	.071	-.416
		3 (正答)		.444	.360		3 (因果)	.286	-.325
		4 (因果)		.037	-.213		4 (正答)	.500	.509
	問3	1 (正答)		.370	.422		1 (否定)	.107	-.010
		2 (因果)		.259	-.143		2 (正答)	.464	.030
		3 (否定)		.148	-.280		3 (因果)	.321	-.146
		4 (対義)		.222	-.305		4 (対義)	.107	-.253
	問4	1 (否定)		.148	-.226		1 (否定)	.071	.261
		2 (対義)		.111	-.178		2 (正答)	.393	.287
		3 (正答)		.481	.380		3 (対義)	.214	-.244
		4 (因果)		.259	-.338		4 (因果)	.321	-.477
冊子2 (29名)	問1	1 (対義)		.103	-.194	冊子4 (28名)	1 (因果)	.357	-.072
		2 (因果)		.448	-.098		2 (対義)	.107	-.152
		3 (正答)		.345	.136		3 (正答)	.393	-.167
		4 (否定)		.103	-.254		4 (否定)	.143	-.071
	問2	1 (正答)		.071	.067		1 (対義)	.179	.312
		2 (対義)		.571	-.004		2 (正答)	.286	-.193
		3 (否定)		.179	-.075		3 (否定)	.214	-.456
		4 (因果)		.179	-.058		4 (因果)	.321	-.015
	問3	1 (否定)		.069	-.258		1 (因果)	.250	-.148
		2 (対義)		.069	-.095		2 (正答)	.357	-.214
		3 (正答)		.414	.569		3 (対義)	.143	-.327
		4 (因果)		.448	-.539		4 (否定)	.250	.230
	問4	1 (否定)		.103	-.019		1 (正答)	.357	.226
		2 (対義)		.103	-.063		2 (否定)	.143	-.475
		3 (因果)		.552	-.467		3 (因果)	.321	-.121
		4 (正答)		.241	.431		4 (対義)	.179	-.107

### 3.3.2 結果

#### 3.3.2.1 項目分析

研究2と同様、第1問は項目単位で、第2問は選択枝単位での項目分析を行った。第1問（共通項目）については、正答率・点双列相関係数ともに適正な範囲に収まっていた。一方、第2問（実験項目）については、一部のテスト冊子において正答選択枝に関する点双列相関係数が負の値を示していたり、錯乱枝に関する点双列相関が正の値を示していた

Table 3.3  
項目分析の結果 (続き)

項目	選択枝	選択率	点双列相関	項目	選択枝	選択率	点双列相関		
冊子 5 (28 名)	問 1	1 (因果)	.357	.061	冊子 8 (30 名)	問 1	1 (正答)	.400	.542
		2 (否定)	.250	-.026			2 (否定)	.200	-.350
		3 (正答)	.214	-.063			3 (対義)	.100	-.291
		4 (対義)	.179	-.248			4 (因果)	.300	-.275
	問 2	1 (正答)	.357	.602		問 2	1 (対義)	.433	-.356
		2 (因果)	.393	-.321			2 (因果)	.133	.030
		3 (対義)	.179	-.341			3 (正答)	.333	.208
		4 (否定)	.071	-.288			4 (否定)	.100	-.160
	問 3	1 (因果)	.222	.019		問 3	1 (対義)	.100	-.168
		2 (否定)	.296	-.621			2 (因果)	.400	-.178
		3 (対義)	.185	-.035			3 (否定)	.133	-.243
		4 (正答)	.296	.438			4 (正答)	.367	.196
	問 4	1 (正答)	.259	.079		問 4	1 (正答)	.400	.058
		2 (否定)	.222	-.592			2 (否定)	.000	NA
		3 (因果)	.148	.032			3 (対義)	.267	.135
		4 (対義)	.370	.172			4 (因果)	.333	-.480
冊子 6 (27 名)	問 1	1 (否定)	.148	-.339	冊子 9 (28 名)	問 1	1 (否定)	.071	.070
		2 (正答)	.556	.261			2 (正答)	.536	.549
		3 (因果)	.074	-.218			3 (因果)	.214	-.487
		4 (対義)	.222	-.177			4 (対義)	.179	-.452
	問 2	1 (正答)	.370	.298		問 2	1 (対義)	.143	-.201
		2 (因果)	.222	.111			2 (因果)	.250	-.364
		3 (否定)	.259	-.429			3 (否定)	.179	-.119
		4 (対義)	.148	-.326			4 (正答)	.429	.340
	問 3	1 (否定)	.148	-.002		問 3	1 (否定)	.143	-.475
		2 (対義)	.111	-.410			2 (因果)	.107	-.204
		3 (因果)	.259	-.138			3 (正答)	.571	.386
		4 (正答)	.481	.109			4 (対義)	.179	-.165
	問 4	1 (否定)	.074	-.474		問 4	1 (正答)	.393	.472
		2 (正答)	.407	.070			2 (因果)	.143	-.154
		3 (因果)	.481	-.033			3 (対義)	.286	-.320
		4 (対義)	.037	-.151			4 (否定)	.179	-.315
冊子 7 (29 名)	問 1	1 (否定)	.138	-.325	冊子 10 (28 名)	問 1	1 (否定)	.321	-.514
		2 (因果)	.172	-.250			2 (正答)	.321	.412
		3 (正答)	.552	.454			3 (因果)	.107	-.237
		4 (対義)	.138	-.325			4 (対義)	.250	.050
	問 2	1 (対義)	.103	-.316		問 2	1 (正答)	.286	.184
		2 (因果)	.103	-.094			2 (因果)	.429	-.026
		3 (否定)	.138	-.382			3 (否定)	.179	-.361
		4 (正答)	.655	.335			4 (対義)	.107	-.155
	問 3	1 (正答)	.483	.347		問 3	1 (否定)	.250	-.414
		2 (対義)	.207	-.314			2 (因果)	.107	.176
		3 (否定)	.207	-.183			3 (正答)	.393	.257
		4 (因果)	.103	-.264			4 (対義)	.250	-.300
	問 4	1 (否定)	.103	-.149		問 4	1 (正答)	.179	.092
		2 (対義)	.069	-.392			2 (否定)	.143	-.234
		3 (因果)	.172	-.251			3 (対義)	.214	-.077
		4 (正答)	.655	.290			4 (因果)	.464	-.015

Table 3.3  
項目分析の結果（続き）

項目	選択枝	選択率	点双列相関	項目	選択枝	選択率	点双列相関		
冊子 11 (27 名)	問 1	1 (正答)	.556	.161	冊子 14 (28 名)	問 1	1 (因果)	.036	-.362
		2 (対義)	.111	-.353			2 (正答)	.679	.641
		3 (否定)	.148	-.156			3 (否定)	.107	-.287
		4 (因果)	.185	-.135			4 (対義)	.179	-.524
	問 2	1 (対義)	.296	.215		問 2	1 (因果)	.071	-.224
		2 (否定)	.185	-.291			2 (否定)	.214	-.353
		3 (因果)	.185	-.123			3 (対義)	.071	-.444
		4 (正答)	.333	.316			4 (正答)	.643	.492
	問 3	1 (対義)	.222	-.162		問 3	1 (正答)	.429	.323
		2 (因果)	.481	.019			2 (因果)	.214	.168
		3 (正答)	.222	.073			3 (対義)	.214	-.565
		4 (否定)	.074	-.276			4 (否定)	.143	-.292
問 4	1 (正答)	.259	.410	問 4	1 (対義)	.107	-.352		
	2 (対義)	.407	-.368		2 (正答)	.464	.375		
	3 (因果)	.222	-.047		3 (否定)	.286	-.334		
	4 (否定)	.111	-.211		4 (因果)	.143	-.083		
冊子 12 (30 名)	問 1	1 (否定)	.033	-.251	冊子 15 (29 名)	問 1	1 (因果)	.345	-.360
		2 (因果)	.200	-.642			2 (否定)	.138	-.051
		3 (対義)	.067	-.125			3 (正答)	.379	.440
		4 (正答)	.700	.555			4 (対義)	.138	-.392
	問 2	1 (因果)	.200	.147		問 2	1 (正答)	.138	-.090
		2 (正答)	.567	.128			2 (対義)	.138	-.312
		3 (否定)	.067	-.205			3 (否定)	.448	.387
		4 (対義)	.167	-.572			4 (因果)	.276	-.297
	問 3	1 (正答)	.533	.487		問 3	1 (否定)	.207	-.330
		2 (対義)	.167	-.245			2 (因果)	.172	.088
		3 (因果)	.267	-.559			3 (対義)	.414	.234
		4 (否定)	.033	-.034			4 (正答)	.207	-.298
問 4	1 (対義)	.200	-.338	問 4	1 (否定)	.103	-.329		
	2 (正答)	.200	-.089		2 (因果)	.103	-.405		
	3 (因果)	.567	.204		3 (対義)	.103	-.181		
	4 (否定)	.033	-.140		4 (正答)	.690	.346		
冊子 13 (30 名)	問 1	1 (因果)	.167	-.304					
		2 (対義)	.033	-.119					
		3 (否定)	.200	.133					
		4 (正答)	.600	-.223					
	問 2	1 (対義)	.267	.082					
		2 (否定)	.267	-.217					
		3 (正答)	.300	-.162					
		4 (因果)	.167	-.090					
	問 3	1 (因果)	.167	-.009					
		2 (正答)	.600	-.122					
		3 (対義)	.100	-.229					
		4 (否定)	.133	-.178					
問 4	1 (正答)	.033	-.116						
	2 (因果)	.200	-.325						
	3 (対義)	.667	.347						
	4 (否定)	.100	-.127						

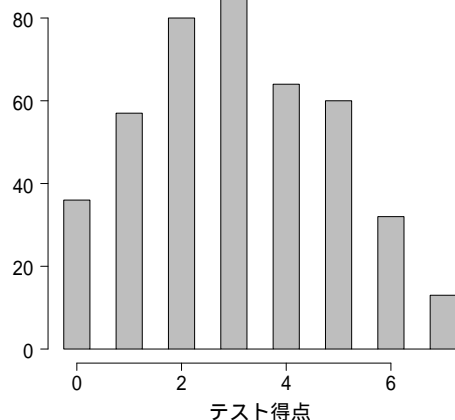


Table 3.4. 冊子ごとの得点の記述統計量

冊子	<i>M</i>	<i>SD</i>	冊子	<i>M</i>	<i>SD</i>
冊子 1	3.407	2.223	冊子 9	3.500	1.991
冊子 2	2.724	1.869	冊子 10	2.464	1.710
冊子 3	3.620	1.916	冊子 11	2.667	1.732
冊子 4	2.929	1.303	冊子 12	3.333	1.729
冊子 5	2.250	1.735	冊子 13	2.833	1.289
冊子 6	3.444	1.847	冊子 14	3.679	2.038
冊子 7	3.862	1.995	冊子 15	2.483	1.573
冊子 8	2.867	1.776			

Figure 3.3. 受検者の得点分布 (研究 4)

りする項目が見られた。例えば冊子 4 では、問 4 を除く 3 項目で正答選択枝に関する点双列相関係数が負の値を示していた。その他、冊子 13 においても 4 項目全体でそのような傾向が見られた。また、冊子 8・問 4 では、否定語錯乱枝を選択する受検者が存在しなかったため、標準偏差が 0 となり、点双列相関係数が算出されなかった。本研究は研究 2 と同様、項目の特性について検討することを目的としていることから、点双列相関係数が低かった項目についても以降の分析に含めることとした。

### 3.3.2.2 受検者の得点分布

受検者のテスト得点（正答数得点）のヒストグラムを Figure 3.3 に示した。受検者の得点の平均は 3.070 点、標準偏差は 1.832 点であった。

テスト冊子ごとにみた受検者のテスト得点について検討を行うため、テスト冊子ごとに得点の平均値と標準偏差を算出した (Table 3.4)。受検者の得点の平均値は、2.00 点～3.70 点の間にあり、テスト冊子に応じて得点の平均値に若干の差が見られた。

### 3.3.2.3 受検者の潜在特性値の推定

研究 2 と同様、共通項目（第 1 問）をアンカーとし、項目応答理論における 1PLM を用いて、受検者の潜在特性値を EAP 推定した。推定にあたり、受検者の潜在特性値  $\theta$  の事前分布は標準正規分布とした。

受検者の潜在特性値の EAP 推定値のヒストグラムを Figure 3.4 に示した。受検者の潜在特性値の推定値の平均は 0.00、標準偏差は 0.78 であった。

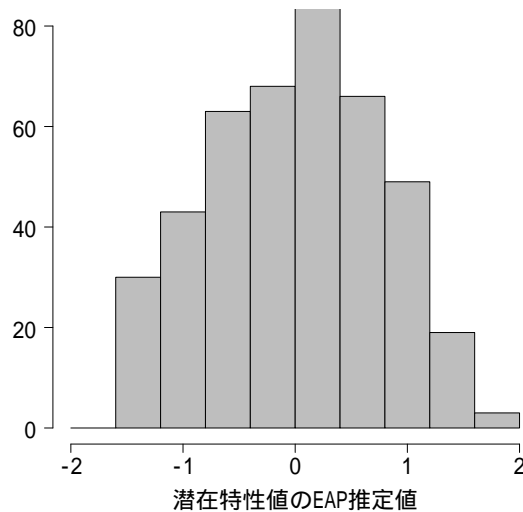


Figure 3.4. 受検者の潜在特性値の EAP 推定値のヒストグラム (研究 4)

### 3.3.2.4 受検者の能力群の構成

研究 2 と同様，受検者の能力水準ごとに選択枝への選択状況に関する検討を行うため，受検者の潜在特性値の EAP 推定値に基づいて能力群を構成した．Kelley (1939) の方法を適用し，受検者の潜在特性値の推定値の分布の 27 パーセンタイル点 (-.498)，73 パーセンタイル点 (.511) を求め，受検者を能力低群 116 名 ( $M = -0.95$ ,  $SD = 0.30$ )，能力中群 198 名 ( $M = 0.01$ ,  $SD = 0.28$ )，能力高群 113 名 ( $M = 0.96$ ,  $SD = 0.33$ ) に分割した．

### 3.3.2.5 錯乱枝の選択率に関する検討

Figure 3.5 に，条件別のトレースラインを示した．一内容を踏まえる設問，同一段落を踏まえる設問では，否定語錯乱枝・対義語錯乱枝の選択率について類似した推移が見られるが，複数段落を踏まえる設問や文章全体を踏まえる設問では，否定語錯乱枝に比べ，対義語錯乱枝の選択率のほうが高いことが見てとれる．また，因果関係の取り違い錯乱枝に関しては，一内容を踏まえる設問や同一段落を踏まえる設問において選択率の高さがうかがえるが，複数段落を踏まえる設問では選択率の低さが示されている．

以下では，キーセンテンスの位置関係に関するカテゴリ (4) × 能力群 (3) × 各選択枝の選択状況 (4) の 3 元分割表を分析対象とし，研究 2 と同様，条件つき局所オッズ比を用いた分析を行った．なお，条件つきオッズ比の統計的推測にあたっては，研究 2 の分析と同様，対数オッズ比の標本分布が漸近的に正規分布に従う性質を利用し，それぞれのオッズ比に対して 95% 信頼区間を算出した．本論文で分析に付した番号を一貫させるため，研

究4のデータを用いての分析を分析4とし、特定の能力群における条件間の選択率の差について検討を行った結果を分析4-1として、特定の条件における能力群間の選択率の差について検討を行った結果を分析4-2として記述した。

#### 分析4-1：特定の能力群における錯乱枝の選択率の条件間差の分析

分析4-1では、能力群ごとにキーセンテンスの位置関係に関するカテゴリ(4)×各選択枝の選択状況(4)を集計した分割表を分析対象とした。この分析では、正答選択枝の選択率に対する各錯乱枝の選択率のオッズを構成し、それを条件間で比較するためのオッズ比について検討を行った。このオッズ比は「一内容を踏まえる」設問を基準とするものであり、他の設問のオッズが「一内容を踏まえる」設問のオッズの何倍であったかを表す。このようなオッズ比を解釈することにより、正答選択枝の選択率を考慮した当該錯乱枝の選択率の条件間での違いについて検討することが可能となる。

分析4-1におけるオッズ比の推定結果をTable 3.5に示した。否定語錯乱枝に関しては、能力群を固定した場合に選択率の異なる条件が特に検出されなかった。対義語錯乱枝に関しては、能力中群・高群において、一内容を踏まえる設問よりも複数段落を踏まえる設問・文章全体を踏まえる設問で選択率が高くなっていた。一方、能力低群においては上記に示した条件間で選択率の差は見られなかった。すなわち、キーセンテンスが比較的広い範囲に散らばっている場合に、能力中群・高群で対義語錯乱枝の選択率が高いことが示されたと言える。因果関係の取り違い錯乱枝に関しては、能力高群において、一内容を踏まえる設問よりも文章全体を踏まえる設問で選択率が高くなっていたが、能力中群においては同一段落を踏まえる設問と比較して、能力低群においては複数段落を踏まえる設問と比較して、一内容を踏まえる設問の選択率が高かった。キーセンテンスが一ヶ所に集約されたような設問では、能力低群・中群において因果関係の取り違い錯乱枝の選択率が高くなることが示された。

#### 分析4-2：特定の条件（設問）における錯乱枝の選択率の群間差の分析

分析4-2では、設問の種類ごとに能力群(3)×各選択枝の選択状況(4)を集計した分割表を分析対象とした。各能力群において正答選択枝の選択率に対する各錯乱枝の選択率のオッズを構成し、それを能力群間で比較するためのオッズ比について検討を行った。このオッズ比は、隣り合う能力群のうち能力水準の高いほうにおけるオッズを基準とするものであり、能力低群におけるオッズが能力中群におけるオッズの何倍であったか、能力中群におけるオッズが能力高群におけるオッズの何倍であったかをそれぞれ表す。このようなオッズ比を解釈することにより、正答選択枝の選択率を考慮した当該錯乱枝の選択率の能力群間での違いについて検討することが可能となる。

分析4-2におけるオッズ比の推定結果をTable 3.6に示した。否定語錯乱枝に関して、能力低群—中群間ではどの条件においても選択率の差があることが示され、能力低群のほう

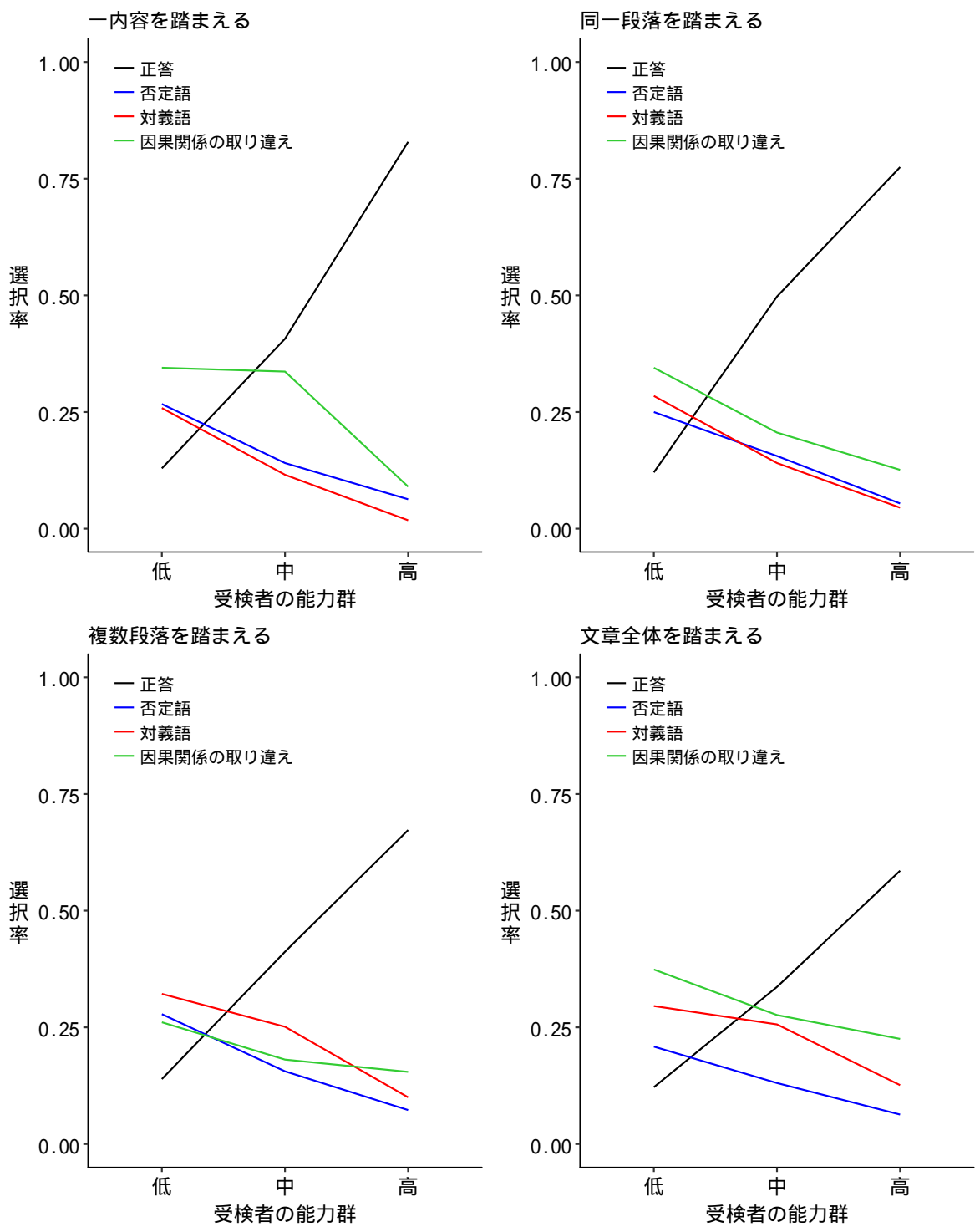


Figure 3.5. 条件ごとにみたトレースライン (研究 4)



Table 3.5  
分析 4-1 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		因果関係の取り違い錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力高群</b>						
同一段落を踏まえる	0.628	[0.219, 1.802]	2.616	[0.494, 13.846]	1.522	[0.669, 3.465]
複数段落を踏まえる	1.066	[0.412, 2.758]	5.921	[1.259, 27.857]	1.830	[0.808, 4.145]
文章全体を踏まえる	1.306	[0.502, 3.401]	11.613	[2.578, 52.314]	3.431	[1.580, 7.452]
<b>能力中群</b>						
同一段落を踏まえる	0.987	[0.546, 1.783]	1.226	[0.670, 2.243]	0.580	[0.354, 0.950]
複数段落を踏まえる	1.093	[0.594, 2.013]	2.232	[1.250, 3.987]	0.781	[0.476, 1.280]
文章全体を踏まえる	0.934	[0.491, 1.777]	2.377	[1.324, 4.268]	1.094	[0.677, 1.767]
<b>能力低群</b>						
同一段落を踏まえる	0.789	[0.343, 1.814]	0.736	[0.320, 1.691]	0.673	[0.308, 1.474]
複数段落を踏まえる	0.925	[0.407, 2.102]	1.025	[0.456, 2.300]	0.372	[0.162, 0.857]
文章全体を踏まえる	0.990	[0.409, 2.397]	1.287	[0.545, 3.040]	0.967	[0.423, 2.212]

Table 3.6  
分析 4-2 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		因果関係の取り違い錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力中群—能力高群</b>						
一内容を踏まえる	3.073	[1.407, 6.715]	13.171	[3.019, 57.466]	6.286	[3.100, 12.746]
同一段落を踏まえる	4.830	[1.920, 12.147]	6.170	[2.303, 16.528]	2.395	[1.253, 4.580]
複数段落を踏まえる	3.153	[1.394, 7.129]	4.965	[2.343, 10.524]	2.682	[1.411, 5.101]
文章全体を踏まえる	2.196	[0.941, 5.129]	2.696	[1.391, 5.224]	2.004	[1.127, 3.565]
<b>能力低群—能力中群</b>						
一内容を踏まえる	5.125	[2.423, 10.841]	6.193	[2.893, 13.258]	3.498	[1.806, 6.776]
同一段落を踏まえる	4.098	[2.048, 8.200]	3.717	[1.861, 7.423]	4.060	[2.125, 7.760]
複数段落を踏まえる	4.337	[2.162, 8.698]	2.843	[1.497, 5.397]	1.667	[0.820, 3.386]
文章全体を踏まえる	5.436	[2.452, 12.054]	3.354	[1.652, 6.810]	3.093	[1.551, 6.169]

が能力中群よりも選択率が高くなっていたが、能力中群—高群では文章全体を踏まえる設問のみにおいて、選択率に能力群間の差が見られなかった。トレースラインを見ても、文章全体を踏まえる設問での否定語錯乱枝の選択率は能力中群—高群で同程度の水準にあり、能力低群においても相対的に低い選択率を示していたことから、文章全体を踏まえる設問での否定語錯乱枝は選択率の低い錯乱枝であった可能性が示唆される。

対義語錯乱枝に関しては、どの条件・どの能力群間においても選択率の差が検出された。特に、一内容を踏まえる設問において大きなオッズ比が確認されたことから、キーセンテンスが一ヶ所に集約されているような設問において、対義語錯乱枝が能力群間の識別に有効であることが示された。

因果関係の取り違い錯乱枝に関しては、多くの条件において能力群間差が検出されたが、複数段落を踏まえる設問において、能力低群と能力中群の選択率の差が見られなかった。

た。トレースラインを確認した場合も、能力低群・能力中群における選択率は同程度であった。また、一内容を踏まえる設問では、能力中群—高群の間で大きなオッズ比が得られた。この結果は、能力中群において一内容を踏まえる設問での因果関係の取り違い錯乱枝の選択率が高いとする分析 4-1 の結果とも整合するものであった。能力中群では、一内容を踏まえる設問において因果関係の取り違い錯乱枝の選択率が能力高群と大きく異なる可能性が示されたと言える。

### 3.3.3 考 察

研究 4 では、研究 3 での整理に基づき、キーセンテンスの位置関係を操作した設問を作成し、否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の選択率について検討を行った。以下では、キーセンテンスの位置関係とそれぞれの選択率の関連について得られた結果を整理し、先行研究や研究 2 での結果と比較しながら、考察を進めることとする。

#### 3.3.3.1 キーセンテンスの位置関係と否定語錯乱枝

否定語錯乱枝に関して、分析 4-1 ではいずれの能力群においても条件間の選択率の差が検出されなかった。このことから、どの能力群においても条件間で選択率は異ならなかったものと考えられる。キーセンテンスが一ヶ所に集約されていた場合も散らばっていた場合も、否定語錯乱枝に対する正誤判断は同じように行われていたことが推察される。

一方、分析 4-2 では多くの条件において能力低群と中群の間、能力中群と高群の間で選択率の差が検出された。例外として、文章全体を踏まえる設問では能力中群と高群の選択率の差が見られなかった。なお、この条件において能力低群・中群との間でのオッズ比は大きく、トレースラインにおいても能力中群よりも能力低群の選択率のほうが高い様子うかがえる。この点を踏まえると、文章全体を踏まえる設問における否定語錯乱枝は、能力低群においてのみ選択率が高く、中群・高群では選択率が低いことが示されたと言える。

研究 2 では、能力低群において文章中の記述なし条件の否定語錯乱枝、能力中群において文章中の記述あり条件の否定語錯乱枝の選択率が高いことを明らかにした。研究 4 の結果を踏まえると、文章中の記述ありの状況の下でも、当該設問におけるキーセンテンスの位置関係の状況によっては、能力中群でも否定語錯乱枝の選択率が低くなりうることが示された。

#### 3.3.3.2 キーセンテンスの位置関係と対義語錯乱枝

対義語錯乱枝に関して、分析 4-1 では能力中群・能力高群の 2 群において、一内容を踏まえる設問に比べ複数の段落を踏まえる設問や文章全体を踏まえる設問での選択率のほうが高かった。このような傾向は、とりわけ能力高群で顕著に見られた。分析 4-2 においても、キーセンテンスの位置関係が集約された範囲にある設問では能力中群—高群の選択率

の差が大きく、広い範囲となる設問では選択率の差が小さくなる傾向が確認された。本研究の結果を踏まえると、対義語錯乱枝は、キーセンテンスの特定が容易な設問下で選択率が低く、キーセンテンスの特定が困難な設問下で選択率が高くなる性質をもつことが示された。ただし、能力低群の受検者はどちらの設問においてもキーセンテンスを特定することが難しく、キーセンテンスの特定が容易な状況においても選択率が高くなるということも明らかとなった。また、対義語錯乱枝は、キーセンテンスが比較的集約されている設問中で用いられた場合に、能力群間の選択率の差が大きくなることも示された。

研究2において、文章中の記述あり条件の対義語錯乱枝に関する検討の結果では、能力低群—中群間で選択率の差が大きくなること、能力高群において対義語錯乱枝の選択率が相対的に高くなることなどが示されている。研究2においては、キーセンテンスの位置関係に関する操作を特に行わずに選択率の検討を行っていたため上記のような結果が得られたが、研究4ではキーセンテンスの位置関係を考慮に入れた設問設定を行ったため、対義語錯乱枝の選択率の差が能力中群—高群間でより大きくなったものと考えられる。

### 3.3.3.3 キーセンテンスの位置関係と因果関係の取り違い錯乱枝

因果関係の取り違い錯乱枝に関して、分析4-1の結果、能力低群では複数段落を踏まえる設問と比べ、一内容を踏まえる設問での選択率が高いことが明らかとなった。能力中群においても類似した結果が見られ、同一段落を踏まえる設問に比べ、一内容を踏まえる設問での選択率が高かった。一方、能力高群では、一内容を踏まえる設問に比べ、文章全体を踏まえる設問において選択率が高かった。能力低群・中群ではキーセンテンスが集約されている設問において因果関係の取り違い錯乱枝の選択率が高くなり、能力高群ではキーセンテンスが広い範囲にある設問において選択率が高くなるという、対照的な結果が得られた。

研究2において、因果関係の取り違い錯乱枝は錯乱枝の理論的分類 (King et al., 2004; Lin et al., 2010) の Level 2 の錯乱枝として位置づけられ、能力中群において文章中の記述あり条件での因果関係の取り違い錯乱枝の選択率が高いという結果が得られていた。その考察の中で、能力中群の受検者が、順接や逆接などの論理的な因果関係や、事象同士の時間的な因果関係の把握を不得手としている可能性について議論したが、研究4においてもこのような考察を支持する結果が得られたと言える。能力低群・中群の受検者は、一内容（隣接する2～3文）同士の論理的な因果関係の理解が十分でないために、順接か逆接かを問わず正しい選択枝であると判断していた可能性がある。能力低群・中群とは異なり、能力高群では、文章全体を踏まえる設問において因果関係の取り違い錯乱枝の選択率が高いという結果が得られた。キーセンテンスが広い範囲にあるような設問は、能力高群においてもキーセンテンスの特定が困難な設問であった可能性が考えられる。なお、能力高群では一内容を踏まえる設問・同一段落を踏まえる設問・複数段落を踏まえる設問での因果関係の

Table 3.7  
研究 5 で用いた文章一覧

冊子	文章の内容	リーダビリティ	語数
共通	国民性	15.30	245
冊子 1	スクールバスのシートベルト	9.14	434
冊子 2	還元主義と創発	14.46	312
冊子 3	自然淘汰	12.40	435
冊子 4	玩具	11.52	357

取り違い錯乱枝の選択率が低かったことから、これらの設問においてはキーセンテンスを十分に特定した上で正誤判断を行っている可能性が考えられる。

### 3.4 研究 5：錯乱枝における使用語と設問タイプが錯乱枝の選択率に及ぼす影響

研究 5 では、キーセンテンスと選択枝の語の重複がある条件とない条件を設定し、設問タイプを考慮に入れた錯乱枝の選択率について検討を行う。

#### 3.4.1 方法

##### 3.4.1.1 受検者

愛知県の大学に通う学生を受検者とした。研究協力に同意のなかった者、すべての項目に解答のなかった者を取り除いた上で、460 名 (男性 140 名, 女性 319 名, 不明 1 名) のデータを分析対象とした。受検者は全員、英語を第二言語とする者であった。

##### 3.4.1.2 材料

インターネット上で公開されている大学の入学試験問題のうち、英語文章読解問題を 5 題用いた。英語文章の選定基準は、研究 2・研究 4 と同様であった。本研究で使用した文章のタイトル・Flesch-Kincaid の公式を用いて算出されたリーダビリティ・語数の一覧を Table 3.7 に示した。

##### 3.4.1.3 設問の作成

本研究では、材料とした英語文章読解問題 5 題のうち共通項目として使用するもの 1 題を除いた 4 題の英語文章に対し、清水 (2005) の定義に沿って、下位レベル設問・上位レベル設問を以下のように作成した。下位レベル設問の作成にあたっては、キーセンテンスを特定できれば選択枝の正誤が判断できるよう、選択枝ごとのキーセンテンスを明確に設定

できる設問として作成した。一方、上位レベル設問では、各段落の主旨に関する問い (e.g., Which of the following statements best summarize each paragraph?) や、文章全体の主旨に関する問い (e.g., What is the main idea of this passage?) を立てた。

下位レベル設問・上位レベル設問ともに、キーセンテンスに基づいた選択枝が複数作成できるよう、問いに対する答えとして複数のキーセンテンスが該当するように作成した。

#### 3.4.1.4 キーセンテンス中の語と錯乱枝中の語の重複に関する実験操作

上記のような設問を作成した後、それぞれの設問に対して選択枝を計4つずつ作成した。選択枝4つの内訳は、正答選択枝・否定語錯乱枝・対義語錯乱枝・因果関係の取り違え錯乱枝であった。キーセンテンス中の語と錯乱枝中の語の重複に関する実験操作を行うため、キーセンテンス中の特定の1語をターゲット語として設定し、ターゲット語として使用する語のみが異なる同内容の錯乱枝を2種類作成した。2種類の錯乱枝のうち、1種類はターゲット語が錯乱枝中でも使用されている条件（重複条件）であり、もう1種類はターゲット語が錯乱枝中で使用されていない、難易度の低い語に置き換えられた条件（非重複条件）である。ターゲット語を操作した錯乱枝の作成手続きは、以下の通りである。

第1に、設問に対する答えとなるようなキーセンテンスを、英語文章中から3つ選定した。これら3つの文は、以降の手続きにおいて、それぞれ否定語錯乱枝・対義語錯乱枝・因果関係の取り違え錯乱枝として作成する際の基になる文であり、この点を念頭に入れてキーセンテンスの選定を行った。例えば、Figure 3.6 に示した錯乱枝の作成にあたっては、以下の3文をキーセンテンスとして選定した。

- The National Highway Traffic Safety Administration has stated that equipping seat belts are not the most effective way to protect schoolchildren.
- Instead, they support the idea of “compartmentalization,” which means sitting in strong, closely spaced seats that have energy-absorbing, high seat backs and having the seats strongly anchored to the floor and frame of the bus.
- The primary argument against adding seat belts is that they would provide very little added protection for impacts of a car collision.

第2に、設問に対するそれぞれのキーセンテンスに含まれる1語をターゲット語として選定した。ターゲット語の選定にあたっては、大学英語教育学会による英単語リスト「JACET8000 英単語」(相澤他, 2005) を参照し、それぞれのキーセンテンス中に含まれるすべての英単語について、使用頻度レベルの確認を行った。その後、キーセンテンス中の語のうち使用頻度の低い1語を選定し、ターゲット語とした。

第3に、ターゲット語をそのまま含むようにして、重複条件の否定語錯乱枝・対義語錯乱枝・因果関係の取り違え錯乱枝を作成した。3種類の錯乱枝は、選定したキーセンテ

スに対して原則 1 対 1 対応するよう作成したが、キーセンテンスと錯乱枝との 1 対 1 対応が難しかった場合には、1 つのキーセンテンスに対して 2 つの錯乱枝を作成したり、2 つのキーセンテンスを部分的に組み合わせて 1 つの錯乱枝を作成することがあった。

第 4 に、重複条件の錯乱枝 3 種類を基にして、非重複条件の否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝をそれぞれ作成した。非重複条件の各錯乱枝を作成する際の方針として、a) ターゲット語の類義語であること、b) JACET8000 の英単語リストにおいてターゲット語よりも使用頻度レベルが高いこと、c) キーセンテンス中で使用されていないことの 3 つを条件に、ターゲット語の置き換え語を検索し、その語をターゲット語の代わりに使用することとした。なお、非重複条件の錯乱枝は、重複条件の錯乱枝に含まれるターゲット語のみを変更したものとして作成し、ターゲット語以外の箇所には変更を加えなかった。ターゲット語の類義語を検索するにあたっては、オンライン版 WordNet Search 3.1(Princeton University, 2015) を使用した。

#### 3.4.1.5 テスト冊子の構成

本研究では計 8 種類のテスト冊子を作成した。それぞれのテスト冊子は、第 1 問・第 2 問の大問 2 問構成であり、第 1 問では英語文章 1 題と多枝選択式問題 3 項目を、第 2 問では英語文章 1 題と多枝選択式問題 4 項目を提示した。なお、第 1 問と第 2 問の英語文章は、互いに異なるものになるようにテスト冊子を作成した。

第 2 問では、いずれの英語文章であっても問 1・問 2 を下位レベル設問とし、問 3・問 4 を上位レベル設問とした。また、第 2 問の間 1～問 4 について、選択枝がすべて重複条件となっている版と、すべて非重複条件となっている版の 2 種類を作成した。最終的に、共通項目として用いた 1 題を除く英語文章 4 題のそれぞれに対して、重複条件の錯乱枝を提示した第 2 問を含むテスト冊子を冊子 1-1 から冊子 4-1 とし、非重複条件の錯乱枝を提示した第 2 問を含むテスト冊子を冊子 1-2 から冊子 4-2 とした。なお、冊子名の数字に関して、ハイフンの前の数字が文章の違いを表しており、ハイフンの後の数字が語の重複に関する条件の違い（重複：1，非重複：2）を表している。

#### 3.4.1.6 手続き

解答データの収集は、2015 年 4 月～9 月までの期間に行われた。受検者には、このテストが受検者自身の能力を測定しようとする目的で行われるものではなく、テスト項目の性質について検討する目的で行われるものであることを伝えた。また、受検者に対する倫理的配慮として、テスト冊子が研究室で厳重に保管されること、研究参加を中断しても受検者自身が不利益を被ることがないことなどを確認した後、受検者の同意を得た。

テストに解答する際、正しいと思う選択枝の丸番号を塗りつぶす形式で解答するよう受検者に教示した。解答時間は 25 分間であった。

下位レベル設問・重複条件

問1 What does the National Highway Traffic Safety Administration state about school buses?

- ① Neither seat belts nor  seats can protect children from car accidents. (否定)
- ② It is necessary for seats in school buses to absorb shocks when a car crash occurs. (正答)
- ③ In case a car  occurs and children were injured, children should wear seat belts. (因果)
- ④  seat belts in school buses is the best way to save children from accidents. (対義)

下位レベル設問・非重複条件

問1 What does the National Highway Traffic Safety Administration state about school buses?

- ① Neither seat belts nor  seats can protect children from car accidents. (否定)
- ② It is necessary for seats in school buses to absorb shocks when a car crash occurs. (正答)
- ③ In case a car  occurs and children were injured, children should wear seat belts. (因果)
- ④  seat belts in school buses is the best way to save children from accidents. (対義)

Figure 3.6. 下位レベル設問 (研究5)

上位レベル設問・重複条件

問3 Which of the following statements best summarizes each paragraph? 6

- ① Paragraph (1) argues that installing strong seats cannot be effective rather than seat belts. (否定)
- ② Paragraph (2) shows that wearing shoulder belts is unsafe for small children. (対義)
- ③ Paragraph (3) states that there are many reasons not to equip seat belts in school buses, other than the matter of protecting children's bodies. (正答)
- ④ Paragraph (4) describes that drivers must be specialized in driving school buses despite of the safety of school buses. (因果)

上位レベル設問・非重複条件

問3 Which of the following statements best summarizes each paragraph? 6

- ① Paragraph (1) argues that setting strong seats cannot be effective rather than seat belts. (否定)
- ② Paragraph (2) shows that wearing shoulder belts is dangerous for small children. (対義)
- ③ Paragraph (3) states that there are many reasons not to equip seat belts in school buses, other than the matter of protecting children's bodies. (正答)
- ④ Paragraph (4) describes that drivers must be professionals in driving school buses despite of the safety of school buses. (因果)

Figure 3.7. 上位レベル設問 (研究5)

## 3.4.2 結 果

### 3.4.2.1 項目分析

これまでの研究と同様、第1問は項目単位で、第2問は選択枝単位での項目分析を行った。第1問において、正答率・点双列相関係数ともに適正な値を示していた。また、第2問においても、正答選択枝に関する点双列相関係数はすべての項目において正の値を示し



Table 3.8  
項目分析の結果

第1問	項目	正答率	点双列相関						
共通項目 (460名)	問1	.537	.378						
	問2	.323	.518						
	問3	.254	.363						
第2問	項目	選択枝	選択率	点双列相関		項目	選択枝	選択率	点双列相関
冊子 1-1 (59名)	問1	1 (否定)	.254	-.409	冊子 1-2 (58名)	問1	1 (否定)	.121	-.161
		2 (正答)	.441	.581			2 (正答)	.414	.420
		3 (因果)	.203	-.157			3 (因果)	.207	-.033
		4 (対義)	.102	-.157			4 (対義)	.259	-.322
	問2	1 (否定)	.153	-.213		問2	1 (否定)	.103	.025
		2 (因果)	.305	-.335			2 (因果)	.397	-.164
		3 (対義)	.136	-.164			3 (対義)	.190	-.186
		4 (正答)	.407	.584			4 (正答)	.310	.315
	問3	1 (否定)	.068	-.027		問3	1 (否定)	.053	.171
		2 (対義)	.288	-.152			2 (対義)	.316	-.407
		3 (正答)	.508	.329			3 (正答)	.439	.229
		4 (因果)	.136	-.260			4 (因果)	.193	.095
問4	1 (因果)	.203	-.131	問4	1 (因果)	.155	-.257		
	2 (対義)	.220	-.306		2 (対義)	.207	-.066		
	3 (正答)	.339	.419		3 (正答)	.379	.310		
	4 (否定)	.237	-.045		4 (否定)	.259	-.070		
冊子 2-1 (57名)	問1	1 (正答)	.193	.480	冊子 2-2 (60名)	問1	1 (正答)	.153	.466
		2 (因果)	.474	-.352			2 (因果)	.339	-.173
		3 (対義)	.228	-.142			3 (対義)	.339	-.233
		4 (否定)	.105	.150			4 (否定)	.169	.065
	問2	1 (対義)	.158	-.262		問2	1 (対義)	.254	.166
		2 (因果)	.193	.098			2 (因果)	.203	-.271
		3 (否定)	.316	-.379			3 (否定)	.220	-.339
		4 (正答)	.333	.494			4 (正答)	.322	.379
	問3	1 (因果)	.158	-.147		問3	1 (因果)	.220	-.060
		2 (正答)	.246	.303			2 (正答)	.305	.328
		3 (否定)	.351	.112			3 (否定)	.254	-.165
		4 (対義)	.246	-.303			4 (対義)	.220	-.130
問4	1 (因果)	.211	-.209	問4	1 (因果)	.136	-.170		
	2 (正答)	.281	.466		2 (正答)	.254	.557		
	3 (否定)	.281	-.297		3 (否定)	.339	-.264		
	4 (対義)	.228	.022		4 (対義)	.271	-.133		

ており、一定程度の識別力があるものと判断された。誤答選択枝に関する点双列相関係数に関しては、一部において0付近の正の値を示すものも見られた。研究2・研究4と同様、本研究の目的が項目の特性について検討を行うことであることを考慮し、誤答選択枝の点双列相関係数が0に近い値をとっていた項目についても以降の分析の対象に含めることとした。

Table 3.8

項目分析の結果（続き）

第2問	項目	選択枝	選択率	点双列相関		項目	選択枝	選択率	点双列相関
冊子 3-1 (60名)	問 1	1 (否定)	.233	.091	冊子 3-2 (57名)	問 1	1 (否定)	.211	-.302
		2 (正答)	.333	.321			2 (正答)	.421	.402
		3 (因果)	.317	-.287			3 (因果)	.228	-.006
		4 (対義)	.117	-.174			4 (対義)	.140	-.210
	問 2	1 (因果)	.217	-.261		問 2	1 (因果)	.228	-.204
		2 (正答)	.300	.437			2 (正答)	.351	.532
		3 (対義)	.317	-.140			3 (対義)	.140	-.271
		4 (否定)	.167	-.074			4 (否定)	.281	-.164
	問 3	1 (正答)	.183	.553		問 3	1 (正答)	.246	.643
		2 (対義)	.350	-.293			2 (対義)	.351	-.240
		3 (否定)	.350	-.212			3 (否定)	.316	-.254
		4 (因果)	.117	.083			4 (因果)	.088	-.157
	問 4	1 (対義)	.217	-.233		問 4	1 (対義)	.211	-.308
		2 (因果)	.183	-.047			2 (因果)	.211	-.147
		3 (否定)	.267	-.145			3 (否定)	.228	-.306
		4 (正答)	.333	.378			4 (正答)	.351	.657
冊子 4-1 (55名)	問 1	1 (対義)	.018	-.076	冊子 4-2 (54名)	問 1	1 (対義)	.148	-.130
		2 (正答)	.509	.575			2 (正答)	.444	.402
		3 (因果)	.291	-.218			3 (因果)	.074	.010
		4 (否定)	.182	-.463			4 (否定)	.333	-.331
	問 2	1 (対義)	.127	-.099		問 2	1 (対義)	.130	-.157
		2 (因果)	.309	-.127			2 (因果)	.352	-.204
		3 (正答)	.255	.255			3 (正答)	.370	.434
		4 (否定)	.309	-.043			4 (否定)	.148	-.167
	問 3	1 (正答)	.418	.401		問 3	1 (正答)	.241	.456
		2 (否定)	.091	-.089			2 (否定)	.204	-.247
		3 (因果)	.382	-.229			3 (因果)	.463	-.213
		4 (対義)	.109	-.195			4 (対義)	.093	.036
	問 4	1 (対義)	.091	-.177		問 4	1 (対義)	.167	-.140
		2 (正答)	.436	.328			2 (正答)	.296	.530
		3 (因果)	.291	-.190			3 (因果)	.333	-.275
		4 (否定)	.182	-.066			4 (否定)	.204	-.150

### 3.4.2.2 受検者の得点分布

受検者のテスト得点のヒストグラムを Figure 3.8 に示した。受検者の得点の平均は 2.454 点、標準偏差は 1.426 点であった。

テスト冊子ごとにみた受検者のテスト得点について検討を行うため、テスト冊子別に得点の平均値と標準偏差を算出した (Table 3.9)。受検者の得点の冊子別平均値は 2 点から 3 点の間に収まっており、テスト冊子による得点の違いは見られないものと判断された。

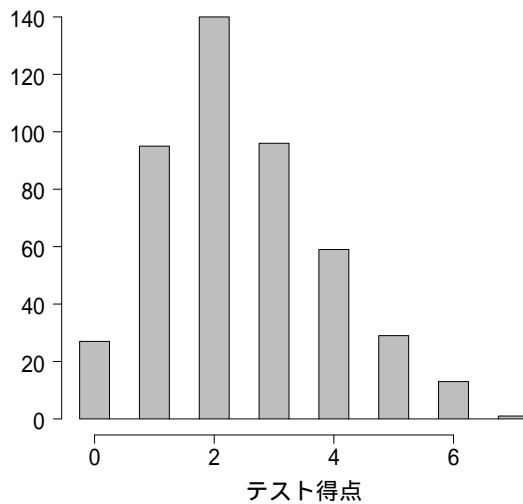


Table 3.9. 冊子ごとの得点の記述統計量

冊子	<i>M</i>	<i>SD</i>
冊子 1-1	2.915	1.611
冊子 1-2	2.741	1.264
冊子 2-1	2.105	1.291
冊子 2-2	2.133	1.186
冊子 3-1	2.283	1.236
冊子 3-2	2.246	1.693
冊子 4-1	2.800	1.445
冊子 4-2	2.426	1.449

Figure 3.8. 受検者の得点分布 (研究 5)

### 3.4.2.3 受検者の潜在特性値の推定

これまでと同様、共通項目（第 1 問）をアンカーとし、項目応答理論における 1PLM を用いて、受検者の潜在特性値を EAP 推定した。推定にあたり、受検者母集団における潜在特性値  $\theta$  の事前分布は標準正規分布とした。

受検者の潜在特性値の EAP 推定値のヒストグラムを Figure 3.9 に示した。受検者の潜在特性値の平均は 0.00、標準偏差は 0.51 であった。

### 3.4.2.4 受検者の能力群の構成

これまでと同様、受検者の能力水準ごとに選択枝への選択状況に関する検討を行うため、受検者の潜在特性値の EAP 推定値に基づいて能力群を構成した。Kelley (1939) の方法を適用し、受検者の潜在特性値の推定値の分布の 27 パーセンタイル点 (-.279)、73 パーセンタイル点 (.275) を求め、受検者を能力低群 129 名 ( $M = -0.58, SD = 0.18$ )、能力中群 207 名 ( $M = -0.03, SD = 0.16$ )、能力高群 124 名 ( $M = 0.66, SD = 0.29$ ) に分割した。

### 3.4.2.5 錯乱枝の選択率に関する検討

Figure 3.10 に、設問タイプ・条件別のトレースラインを示した。全体的に見て、重複条件・非重複条件を問わず、上位レベル設問における錯乱枝は種類による違い（否定語・対義語・因果関係の取り違い）が小さかったのに対して、下位レベル設問における錯乱枝は選択率の差が顕著に見られた。この結果から、否定語や対義語、因果関係の取り違いなどを含んだ錯乱枝は、下位レベル設問において選択率の差が顕著になる可能性が示唆された。

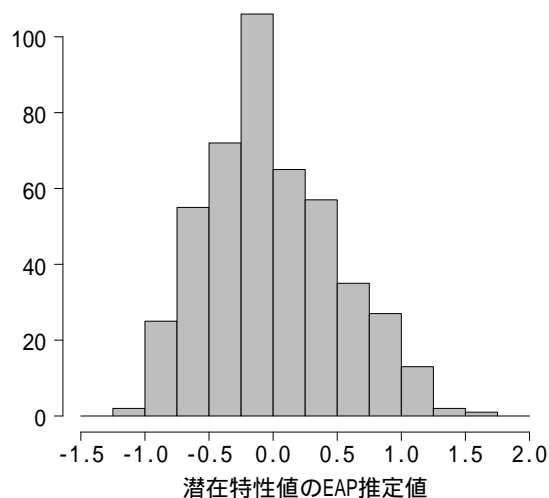


Figure 3.9. 受検者の潜在特性値の EAP 推定値のヒストグラム

下位レベル設問・重複条件では，錯乱枝の種類の違いが選択率に大きな差をもたらしており，因果関係の取り違え錯乱枝の選択率が他の錯乱枝に比べ高くなるなど，研究4におけるトレースラインと一致する傾向が確認された。

以下では，能力群(3)×設問タイプ(4)×語の重複に関する条件(2)×各選択枝の選択状況(4)の4元分割表を分析対象とし，これまでの研究と同様，条件つき局所オッズ比を用いた分析を行った。なお，条件付き局所オッズ比の統計的推測にあたっては，対数オッズ比の標本分布が漸近的に正規分布に従う性質を利用し，それぞれのオッズ比に対して95%信頼区間を算出した。研究5のデータを用いての分析を分析5とし，能力群・設問タイプを固定した場合における語の重複の条件間の選択率の差について検討を行った結果を分析5-1，設問タイプ・語の重複の条件を固定した場合における能力群間の選択率の差について検討を行った結果を分析5-2とした。

#### 分析 5-1：特定の能力群における錯乱枝の選択率の設問タイプ間・条件間差の検討

分析5-1では，能力群・設問タイプごとに語の重複に関する条件(2)×各選択枝の選択状況(4)を集計した分割表を分析対象とした。この分析では，重複条件・非重複条件のそれぞれについて正答選択枝の選択率に対する各錯乱枝の選択率のオッズを構成し，オッズ比の分母に非重複条件のオッズ，分子に重複条件のオッズを置いて，重複条件におけるオッズが非重複条件におけるオッズの何倍であるかについて検討を行う。オッズ比が1より大きかった場合には重複条件の錯乱枝のほうが選択されやすく，オッズ比が1より小さかった場合には非重複条件の錯乱枝のほうが選択されやすいことを表す。

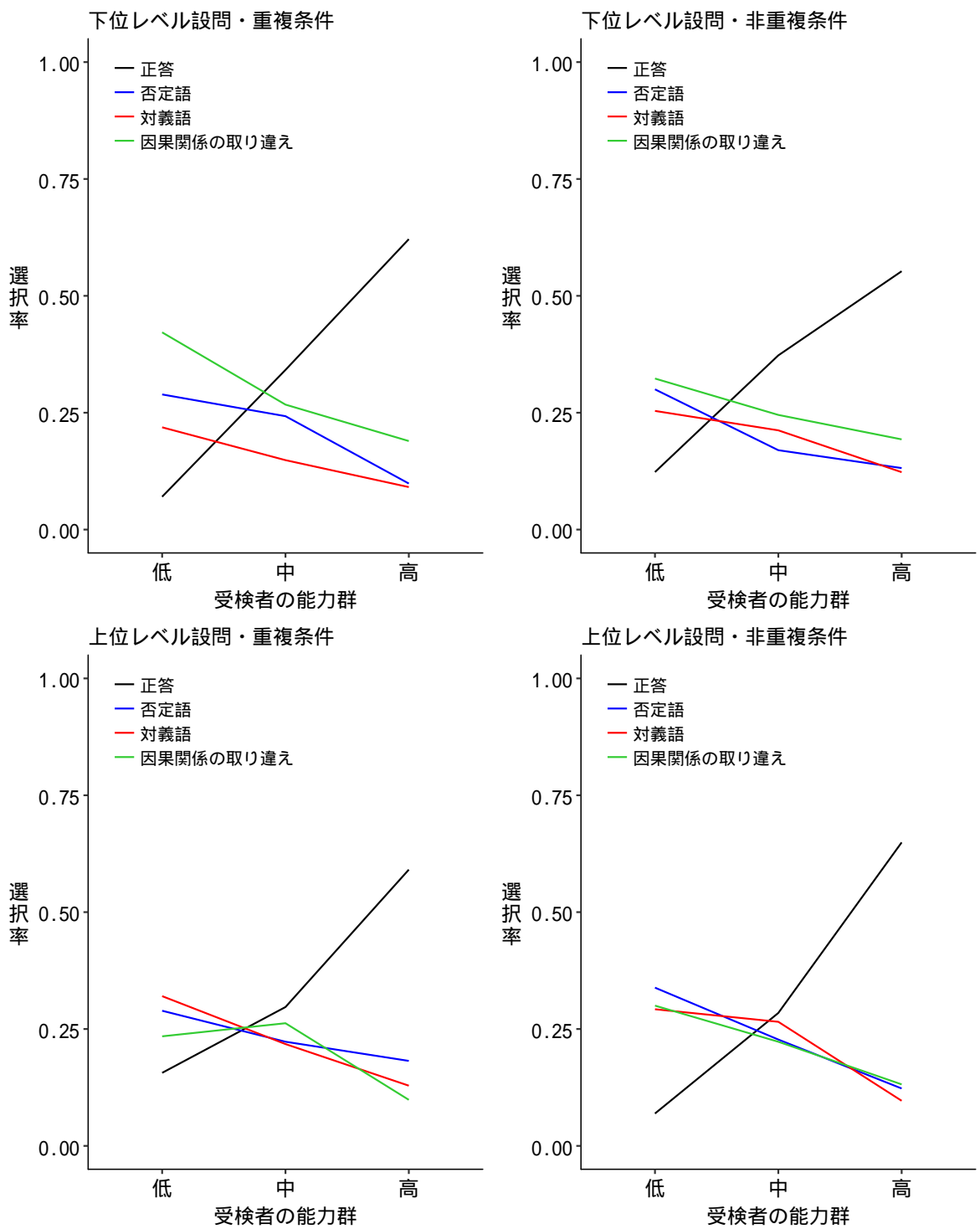


Figure 3.10. 設問タイプ・条件ごとにみたトレースライン (研究 5)

Table 3.10  
分析 5-1 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		因果関係の取り違い錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力高群</b>						
下位レベル設問	0.666	[0.296, 1.500]	0.659	[0.285, 1.522]	0.873	[0.451, 1.690]
上位レベル設問	1.626	[0.782, 3.381]	1.466	[0.644, 3.337]	0.822	[0.367, 1.844]
<b>能力中群</b>						
下位レベル設問	1.558	[0.910, 2.669]	0.763	[0.434, 1.341]	1.189	[0.722, 1.959]
上位レベル設問	0.938	[0.545, 1.612]	0.786	[0.461, 1.339]	1.128	[0.663, 1.918]
<b>能力低群</b>						
下位レベル設問	1.687	[0.664, 4.285]	1.508	[0.578, 3.937]	2.286	[0.919, 5.683]
上位レベル設問	0.378	[0.154, 0.931]	0.486	[0.197, 1.197]	0.346	[0.138, 0.868]

分析 5-1 におけるオッズ比の推定結果を Table 3.10 に示した。全体として、重複条件と非重複条件の間のオッズ比は 1 付近の値をとっており、いずれの能力群においても両条件間で選択率の差が見られなかったと言える。例外として、能力低群では上位レベル設問で否定語錯乱枝・因果関係の取り違い錯乱枝のオッズ比が小さい値を示していた。上位レベル設問では、能力低群において非重複条件の否定語錯乱枝・因果関係の取り違い錯乱枝の選択率が高いことが示された。

その他の条件においては信頼区間に 1 が含まれていたが、オッズ比そのものを解釈した場合には、次のようなことが確認できる。まず、否定語錯乱枝に関して、能力低群では下位レベル設問・重複条件における選択率のほうが高かったことがわかる。同様の傾向は能力中群でも確認できる。このことを踏まえると、能力低群・中群では、下位レベル設問において重複条件の否定語錯乱枝の選択率が高い可能性が示唆された。一方、能力高群においては、下位レベル設問において非重複条件のオッズのほうがやや大きく、上位レベル設問では重複条件のオッズのほうが大きい傾向が見られた。この結果から能力高群では、下位レベル設問において非重複条件、上位レベル設問において重複条件の否定語錯乱枝の選択率が高くなる可能性が示された。

次に、対義語錯乱枝に関して、能力低群では下位レベル設問・重複条件における選択率のほうが高かったことがわかる。能力中群では、いずれの設問タイプにおいても重複条件と非重複条件との間で特に大きな差は見られなかったが、能力高群では下位レベル設問・非重複条件の選択率が高かった。

因果関係の取り違い錯乱枝に関して、能力低群では下位レベル設問・重複条件における選択率が高くなっていったが、能力中群・高群ではいずれの設問タイプにおいても条件間の選択率の差は検出されなかった。

#### 分析 5-2 : 特定の設問タイプ・条件における錯乱枝の選択率の群間差の検討

Table 3.11  
分析 5-2 におけるオッズ比の推定結果

条件	否定語錯乱枝		対義語錯乱枝		因果関係の取り違い錯乱枝	
	OR	95%CI	OR	95%CI	OR	95%CI
<b>能力中群—能力高群</b>						
下位レベル設問・重複	4.479	[2.246, 8.933]	2.971	[1.415, 6.240]	2.567	[1.449, 4.549]
下位レベル設問・非重複	1.914	[0.963, 3.806]	2.563	[1.292, 5.085]	1.885	[1.036, 3.429]
上位レベル設問・重複	2.438	[1.339, 4.437]	3.365	[1.751, 6.465]	5.300	[2.648, 10.606]
上位レベル設問・非重複	4.229	[2.130, 8.395]	6.279	[3.024, 13.037]	3.864	[1.970, 7.579]
<b>能力低群—能力中群</b>						
下位レベル設問・重複	5.789	[2.562, 13.083]	7.156	[3.014, 16.990]	7.667	[3.478, 16.900]
下位レベル設問・非重複	5.349	[2.649, 10.802]	3.621	[1.797, 7.294]	3.988	[2.033, 7.823]
上位レベル設問・重複	2.467	[1.266, 4.807]	2.795	[1.443, 5.415]	1.698	[0.864, 3.338]
上位レベル設問・非重複	6.111	[2.715, 13.755]	4.524	[2.007, 10.197]	5.532	[2.438, 12.550]

分析 5-2 では、設問タイプ・語の重複に関する条件ごとに能力群 (3) × 各選択枝の選択状況 (4) を集計した分割表を分析対象とした。この分析では、当該の設問・条件での正答選択枝の選択率に対する各錯乱枝の選択率のオッズを算出し、それを能力群間で比較するオッズ比について検討を行った。分析 4-2 と同様、このオッズ比は、隣り合う能力群のうち能力水準の高いほうのオッズを基準としたものであり、能力低群におけるオッズが能力中群の何倍であるか、能力中群におけるオッズが能力高群の何倍であるかを表している。このようなオッズ比を構成することにより、当該の設問・条件において正答選択枝の選択率を考慮したときの各錯乱枝の選択率が能力群間で異なるかどうかについて検討した。

分析 5-2 におけるオッズ比の推定結果を Table 3.11 に示した。多くの設問タイプ・条件において、能力群間の選択率の差が見られた。すべての条件を通じて、能力高群よりも中群のほうが、中群よりも低群のほうが選択率が高かったことがうかがえる。特に、下位レベル設問における重複条件では、能力低群と中群の間で選択率の差が大きいことが確認された。ただし、一部の条件において能力群間の選択率差が検出されなかったものもあった。具体的には、否定語錯乱枝に関して、下位レベル設問・非重複条件では能力中群と高群との間に選択率の差が見られなかった。また、因果関係の取り違い錯乱枝に関して、上位レベル設問・重複条件では能力低群と中群との間に選択率の差が見られなかった。

### 3.4.3 考 察

本研究では、キーセンテンス中に含まれる語が錯乱枝中でも含まれている場合と、より難易度の低い語に置き換えた場合で錯乱枝の選択率が異なるかどうかについて、設問タイプの違いを考慮に入れて能力群別に検討を行うことが目的であった。以下では、設問タイプ・語の重複とそれぞれの錯乱枝の選択率との関連について得られた知見を整理し、先行研究での結果と比較しながら、考察を進めることとする。

### 3.4.3.1 下位レベル設問

清水 (2005) では、文章中に明示されている情報についての理解を問う設問を下位レベル設問としており、本研究でもこの定義に沿って設問を作成し、錯乱枝に関する検討を行った。その結果、能力低群では非重複条件よりも重複条件の否定語錯乱枝・因果関係の取り違い錯乱枝の選択率がやや高く、能力中群では重複条件の否定語錯乱枝の選択率がわずかに高かった。能力高群では、重複条件に比べ非重複条件の否定語錯乱枝の選択率が高かった。下位レベル設問の結果を総括すると、能力低群・能力中群の受検者が語の重複を含む錯乱枝を選び、能力高群の受検者がターゲット語を置き換えた錯乱枝を選んでいった。

下位レベル設問では、キーセンテンスを特定した上で選択枝の正誤を判断することが求められる。先行研究では、読解能力の低い受検者であっても語の重複を利用して正誤判断を行うことが示されている (Buck et al., 1997)。ただし、読解能力の低い受検者は、キーセンテンスと同一の語が錯乱枝中に含まれているだけで、錯乱枝に記述された内容を正しいと判断している可能性が考えられる。本研究の結果はこの想定を支持し、能力低群の受検者において、重複条件における否定語錯乱枝・因果関係の取り違い錯乱枝の選択率が高かった。このことを踏まえると、読解能力の低い受検者がキーセンテンスと錯乱枝の語の重複のみに基づいて、これらの錯乱枝を正しいと判断していたことが推察される。さらに、分析 5-2 では、重複条件において能力群間の差が大きくなり、特に能力低群と中群の間の選択率の差が大きくなるという結果が得られた。この結果も、上記の想定を支持するものであると考えられる。

正答選択枝の選択率に対して語の重複の効果を検討した Freedle & Kostin (1993) では、下位レベル設問における語の重複の効果が能力群を問わず検出されていたが、研究 5 では錯乱枝の選択率に対する語の重複の効果が能力低群・能力中群で検出された。先行研究と研究 5 の結果を統合すると、読解能力の低い受検者は正答選択枝がどれであるか見当がつかないため、すべての選択枝に対して語の重複の手がかりを用いてキーセンテンスを特定しようとした結果、語の重複のある正答選択枝・錯乱枝ともに選択率が高まることが考えられる。一方で、読解能力の高い受検者は、正答選択枝に絞って語の重複の手がかりを用いることが想定される。

### 3.4.3.2 上位レベル設問

清水 (2005) では、文章全体の構造やテーマの理解などを問う設問を上位レベル設問としており、本研究でもこの定義に沿って設問を作成し、錯乱枝に関する検討を行った。その結果、能力低群では重複条件に比べ非重複条件の否定語錯乱枝・因果関係の取り違い錯乱枝の選択率が高いという結果が得られた。さらに分析 5-2 では、上位レベル設問・非重複条件における能力低群の選択率が能力中群に比べても高いという結果が得られた。上位



レベル設問では、段落や文章全体の主旨を把握した上で、設問で指示された段落や文章全体の主旨が適切に反映された選択枝を選ぶことが求められる。このような設問では、一見キーセンテンスの特定が不要であるように思えるが、実際にはキーセンテンスの特定という認知過程が必要となる。先行研究では、使用頻度の高い語を用いた選択枝の選択率が高くなる (Drum et al., 1981) ことが知られている。先行研究および研究 5 の結果を踏まえると、使用頻度の高い語に置き換えることにより、錯乱枝に記述された内容がわかりやすくなり、読解能力の低い受検者はキーセンテンスの特定という認知過程を経ずに選択枝の正誤判断をしていた可能性が考えられる。

なお、非重複条件の否定語錯乱枝や対義語錯乱枝においては、能力中群—高群間の選択率の差が大きかった。能力中群においても能力低群と同様、否定語や対義語を含む錯乱枝の内容が易しい語で書かれていた場合に、その錯乱枝を正しいと判断する傾向がある可能性が示唆される。一方で、因果関係の取り違い錯乱枝に関しては、重複条件において能力中群と能力高群との間で選択率の差が大きくなる傾向が確認された。能力中群の受検者にとっても因果関係の判断が難しい可能性があることはこれまで述べてきたとおりであるが、キーセンテンス中の語が錯乱枝中にも含まれている場合に、因果関係に対する判断がより難しくなる可能性が示唆された。

### 3.5 第 3 章のまとめ

第 3 章では、文章読解問題解答時の情報処理モデル (Embretson & Wetzel, 1987) に基づき、選択枝の内容を踏まえて英語文章中のキーセンテンスを特定する過程や、キーセンテンスと選択枝を比較対照する過程に着目した検討を行った。このような検討の背景には、選択枝の正誤判断の結果としての各錯乱枝の選択率が、キーセンテンスを特定したり、キーセンテンスと選択枝を比較したりする過程の影響を受けている可能性があり、それらの認知過程はテストにおける設問設定が規定している点を考慮したことが挙げられる。第 3 章では、1) キーセンテンスの特定の容易さを、キーセンテンスの位置関係という観点から設問で操作した場合の錯乱枝の選択率、2) キーセンテンスと選択枝の比較対照の容易さを、選択枝中の使用語という観点から設問で操作した場合の錯乱枝の選択率、の 2 点から検討を行った。以下では、第 3 章での研究課題 2 つに沿って得られた結果を整理しながら、項目作成に対する示唆を述べることとする。

第 1 の研究課題「設問においてキーセンテンスの特定の容易さを操作した場合」に関しては、対義語錯乱枝・因果関係の取り違い錯乱枝において特徴が見られ、キーセンテンスが比較的一ヶ所に集約されるよう作成した設問では、能力群間の選択率に大きな差が確認された一方で、キーセンテンスが文章全体に点在するよう作成した設問では、能力高群においてもこれらの錯乱枝の選択率が高くなり、その結果として能力群間の選択率の差が小

さくることが示された。これらの知見を踏まえた項目作成の例として、大規模テストで識別力の高い英語文章読解問題を作成する必要がある状況では、キーセンテンスが一ヶ所に集約されるように設問を作成し、対義語錯乱枝や因果関係の取り違い錯乱枝を用意することが望まれる。一方、難易度の高い英語文章読解問題を作成する必要がある状況では、キーセンテンスが英語文章中の広い範囲に散らばるように設問を作成し、対義語錯乱枝や因果関係の取り違い錯乱枝を用いることが望まれる。この場合、これらの錯乱枝に関する識別力はそれほど高くない点に注意が必要である。

第2の研究課題「設問においてキーセンテンスと選択枝の比較対照の容易さを操作した場合」に関しては、下位レベル設問と上位レベル設問で異なる特徴が確認され、下位レベル設問では能力水準の低い受検者において重複条件の錯乱枝の選択率が高く、能力水準の高い受検者において非重複条件の錯乱枝の選択率が高いという結果が示されたが、上位レベル設問では下位レベル設問とは逆の結果が得られた。このような結果に対応して、下位レベル設問では重複条件において能力群間でのオッズ比が大きく、上位レベル設問では非重複条件において能力群間でのオッズ比が大きくなっていった。これらの知見は、キーセンテンスと選択枝を見比べて正誤判断する際の受検者の解答プロセスを示唆しているものと考えられる。キーセンテンスの特定を求める設問を作成する必要がある場合には、難易度の高い語であったとしても、キーセンテンス中の語を錯乱枝中に含めたほうが、能力群間での選択率の差が大きくなるため、結果として識別力の高いテスト項目になることが期待される。一方で、段落・文章全体の構造の理解を求めるような設問を作成する必要がある場合には、難易度の高い語を易しい語に置き換えた錯乱枝を作成したほうが、能力群間での選択率の差が大きくなるため、識別力の高いテスト項目になることが期待される。

なお、第1の研究課題で検討した「キーセンテンスが英語文章中の広い範囲に散らばるような設問」と、第2の研究課題で検討した「段落・文章全体の構造の理解を求めるような設問」は、いずれも英語文章全体を見渡した内容の理解を受検者に求めているという点で共通している。そのため、キーセンテンスが英語文章中の広い範囲に散らばる設問において、能力群間での錯乱枝の選択率の差が小さいという結果と、段落・文章全体の構造の理解を求めるような設問において、能力群間での錯乱枝の選択率の差が大きくなる条件があるとする結果は、一見競合するように見えるかもしれない。しかし、これらの結果の違いは、英語文章中の広い範囲の内容の読解を求める項目を作成する際の方針を、以下のような形で示しているものと考えられる。

研究4における「キーセンテンスが英語文章中の広い範囲に散らばる設問」では、英語文章中の広い範囲からキーセンテンスを実際に抽出し、その内容に基づいて錯乱枝を作成した。この設問で能力高群における選択率が他の設問と比べて高くなっていたことを踏まえると、設問幹・選択枝を通じてキーセンテンスを特定するための情報が十分に与えられていなかったため、能力高群の受検者にとってもキーセンテンスの特定が困難であった可

能性が考えられる。この点を解決する一つの策として、研究5のように段落や文章全体の主旨を問う設問として設定し、それぞれの錯乱枝がどの位置にあるキーセンテンスの特定を求めているか明確にすることが挙げられる。このような設問設定としたことで、キーセンテンスの特定に必要な情報が十分に受検者に与えられ、能力高群の受検者の多くが錯乱枝を選択しない設問になったことが考えられる。上記のことを踏まえると、英語文章中の広い範囲の内容の読解を求める項目を作成する際、能力群間の選択率の差が大きくなるために、受検者がキーセンテンスを特定するために必要な情報を含めるような設問設定を行うことの必要性が示唆されたと言える。

第3章における2つの研究課題に対して得られた知見はいずれも、テストの仕様に対応した設問の作成、錯乱枝の配置が必要であることを実証的に示している。設問の設定にもきめの細かい配慮を施した上で錯乱枝を作成する必要があるという点は、実際の項目作成の現場においても考慮すべき点であると考えられる。



## 第4章

# 英語文章の要点を把握する能力を測る問題の錯乱枝開発

### 4.1 研究の背景

前章までは、選択枝の内容に関連する文を特定するという局所的な情報処理の側面を取り上げ、錯乱枝に関する検討を行ってきた。本章では、局所的な情報処理から大域的な情報処理に焦点を移し、段落や文章全体の要点を把握する側面に着目して、錯乱枝に関する検討を行う<sup>3</sup>。

このような読解能力を取り上げる理由として、要点の把握が読解と密接に結びついた側面であることが挙げられる。Karbalaee & Rajyashree (2010) は、英語文章の要約プロセスを学ぶプログラムに参加した学習者の TOEFL の得点が、プログラムの参加前後で上昇していたことを報告している。この研究で使用された TOEFL の問題には、要点把握の側面だけでなく、局所的な情報処理や筆者の意図の推測などの側面の読解問題も含まれていた。また、その他の先行研究においても、要点を把握するような読み方が英語文章読解と密接に関連していることが示されている (e.g., Oded & Walters, 2001; Zhao & Hirvela, 2015)。これらの点を踏まえると、英語文章読解問題のバリエーションのひとつとして、要点把握の側面を測る読解問題を取り上げて検討を行うことには意義があると考えられる。

英語文章の要点の把握を取り上げるにあたって、前章までとは異なる側面の英語文章読解能力であるため、これまでに扱ってきた誤答の特徴 (e.g., 否定語・対義語・因果関係の取り違い) を引き続き用いることが適切であるかどうかは定かでない。そこで本章では、英語文章の要点把握に深い関わりがあると思われる、文章要約についての認知的心理学的知見を踏まえ、要点の把握を行う際に特有の誤答を改めて抽出する。

第1章でも述べたとおり、Kintsch (1998) の構築—統合モデルでは、文中に含まれる命題表象をマイクロ構造として構築した後、文章全体の命題表象をマクロ構造として構築するこ

---

<sup>3</sup> 本章の一部は、寺尾・石井 (印刷中) を加筆・再構成したものである。

とが仮定されている。マイクロ構造として構築された個々の命題をまとめるためには、重要な情報を取捨選択する（削除）、具体的な表現をより上位概念の語に置き換える（一般化）、段落を代表する文（トピック・センテンス）を産出する（統合）などの過程を経る (Kintsch & van Dijk, 1978)。本章では、情報の取捨選択を行う削除過程、具体性の高い表現から上位概念の語へと置き換えを行う一般化過程、トピック・センテンスを産出する統合過程のそれぞれに着目し、各過程における誤答の特徴を把握した上で、その特徴を含んだ錯乱枝について検討を行う。

英語文章の要点把握の際の誤答の特徴を把握する上で、英語を第二言語とする学習者 (L2 学習者) の要約文の特徴を明らかにした研究知見は、重要な役割を果たすものと考えられる。第二言語習得に関する先行研究では、L2 学習者の要約文に大きく 2 つの特徴があることが明らかにされている (e.g., Johns & Mayes, 1990; Keck, 2006; Kim, 2001)。

L2 学習者の要約文の特徴の 1 つ目として、複写一削除方略 (copy-delete strategy) の使用が挙げられる (Brown & Day, 1983; Johns & Mayes, 1990; Keck, 2006)。複写一削除方略とは、英語文章中で利用可能な表現があれば、そのまま複写するか微修正のみを加えて要約文に含めるという方略のことである。Johns & Mayes (1990) は、このような特徴が能力水準の低い L2 学習者に見られ、能力水準が高くなるにつれて要約文中に自らの言葉で言い換えた表現が増加することを報告している。複写一削除方略の使用は、情報の取捨選択を必要とする削除過程では適切である一方で、具体的な表現をより上位概念の語に置き換える一般化過程では不適切な場合もあると考えられる。例えば、複写一削除方略を使用した学習者は文章中の記述をそのまま利用し、具体的な表現ばかりの要約文を作成するのに対して、複写一削除方略を使用しない学習者は、文章中の表現をできるだけ言い換え、より文章全体の主張に即した要約文を作成することが想定される。

L2 学習者の要約文の特徴の 2 つ目として、複数の命題をまとめた表現が少ないということが挙げられる (Johns & Mayes, 1990; Kim, 2001)。Kim (2001) では、韓国人の L2 学習者における要約文の中に、文章中の 2 つ以上の命題を 1 つに置き換えた記述が少なかったことが報告されている。また、Johns & Mayes (1990) においては、能力水準の高低を問わずこのような特徴が見られたことを報告している。複数の命題を 1 つの文に反映させることは、主にトピック・センテンスを産出する統合過程でより顕著に見られるものと考えられる。どの能力水準にある L2 学習者も、文章中の命題の統合を苦手としている可能性が示唆される。

以上のように見ると、英語を第二言語とする受検者の要約文の特徴を捉えることにより、能力水準別の受検者の要約パターンを浮き彫りすることができ、英語文章の要点を把握する能力を測定する多枝選択式問題の錯乱枝の作成に活用することが可能になるものと考えられる。そこで第 4 章では、英語文章の要点を把握する能力に焦点を当て、受検者の要約文における解答パターンの情報に基づいて錯乱枝を作成した場合の選択率について、

検討を行う。

以上の点を踏まえ、本章では英語文章の要点把握能力を測る問題の錯乱枝に関する知見を得るため、次の2つの研究課題を設定した。第1の研究課題は、記述式の要約問題を用いて受検者の解答パターンを把握し、要点把握に特有の誤答の特徴を抽出することである(研究6)。第二言語習得の領域では、英語文章を要約する際の特徴について多くの検討が行われているが(e.g., Johns & Mayes, 1990; Keck, 2006; Kim, 2001)、あくまで第二言語習得に関する研究として行われたものであるため、誤答に関する情報は十分に得られていない。英語文章の要点把握の側面を取り上げ、錯乱枝の作成に活用するためには、教育測定学的な観点から誤答についての詳細な検討を追加する必要があると考えられる。

第2の研究課題は、第1の研究課題の結果を踏まえて錯乱枝を作成し、能力群別の選択率について検討することである(研究7)。錯乱枝について検討を行った先行研究のうち、英語文章の要点を把握する能力を題材としたものは見当たらない。本章は、前章までの知見に、大域的な情報処理の側面の読解能力を測る際に有効となる錯乱枝についての知見を追加する役割を果たすものであり、工夫を施した錯乱枝が英語文章読解問題作成のバリエーションを増やすことに寄与する可能性がある。

## 4.2 研究6：英語文章要約問題における典型的な誤答パターンの検討

研究6では、本章の第1の研究課題として挙げた「英語文章要約における解答パターンの把握」を行うため、記述式の英語文章要約問題を用いて受検者の誤答の特徴を明らかにする。

### 4.2.1 方法

#### 4.2.1.1 受検者

国立大学に通う大学生60名(男性30名、女性30名)がテストへの解答に参加した。ただし、2名の受検者が全問無解答であったため、分析には58名の解答データを使用した。国立大学に通う大学生を受検者としたのは、記述式の要約問題への解答にあたって、一定程度のライティングスキルが必要となることを考慮したためである。本研究でリクルートされた受検者は、大学入試において記述・論述式の個別学力試験を受検した経験をもつ大学生であったことから、一定程度のライティングスキルがあり、本研究においても要約文の作成を問題なく行うことができると判断される。

Table 4.1  
研究 6 で用いた文章一覧

冊子	テーマ	リーダビリティ	語数
1	休暇	8.63	799
2	金銭とモチベーション	9.39	849
3	産業革命の影響	10.32	702

#### 4.2.1.2 材 料

オンラインで公開されている大学の入学試験問題のうち、英語文章読解問題を 4 題使用した。4 題の英語文章読解問題のうち、1 題は全体的な英語文章読解能力を測定するための問題として入学試験のまま使用し、残り 3 題は要約問題用として、本研究の目的に合うよう適宜修正を行い使用した。要約問題の作成の詳細は、次節にて述べる。

先行研究では、文章のさまざまな特徴 (e.g., 文章の長さ, ジャンル, 複雑さ) が要約文のさまざまな特徴に影響を及ぼすことが知られている (Hidi & Anderson, 1986)。本研究では、テストへの解答時間に限りがあり、極端に長い文章を提示することが難しいことから、文章の語数のある程度揃え、700~850 語程度の文章 3 題を選定した。一方、実際の英語文章読解テストにおいてさまざまなジャンルの文章が用いられている実状に照らして、文章のジャンルを制限するような手続きはとらなかった。文章の複雑さに関して、Hidi & Anderson (1986) ではさまざまな特徴を挙げており、そのひとつに文構造の複雑さを挙げている。本研究では、前章までの研究でも用いてきたリーダビリティの指標が文構造の複雑さを一定程度反映する指標と考え、米国の学年にして 9 年生 (わが国の中学 3 年生に相当) 程度の文章を選定した。本研究で使用した文章のテーマ・リーダビリティ・語数の一覧を Table 4.1 に示した。

#### 4.2.1.3 英語文章要約問題の作成

本研究では、英語文章 1 題につき要約問題 4 項目を作成した。要約問題 4 項目の内訳は、設問で指定された段落を要約する問題 (以下、指定段落要約問題) 3 項目、文章全体を要約する問題 (以下、文章全体要約問題) 1 項目であった。これらの要約問題を英語文章 3 題分、計 12 項目を作成した。

指定段落要約問題では、各認知過程における受検者の誤答の特徴を把握するという目的を達成するため、削除・一般化・統合のうちいずれかを各項目に割り当てた (問 1 は削除, 問 2 は一般化, など)。ただし、1 つの段落中に含まれている命題数は限られているため、段落の要約では顕在化しにくい特徴もあるものと考えられる。そこで、文章全体の要約では削除・一般化・統合のプロセスがすべて関与することを仮定し、文章全体要約問題を設



定した。文章全体要約問題を通して、指定段落要約問題では抽出することの難しい誤答の特徴を把握することが可能になる。

指定段落要約問題では、割り当てた認知プロセスが要約時に特に重要となるような段落を選定し、設問の幹において「第2段落を要約してください」のように教示した。要約を求める段落の選定にあたっては、それぞれの認知プロセスの特徴を考慮しながら、次のような基準で選定した。削除を割り当てた要約問題（以下、指定段落要約問題・削除）では、類似する内容が複数回述べてられている、周辺的な内容を多く含んでいるなど、要約文に含めるべき内容の取舍選択が必要となるような段落を選定した。一般化を割り当てた要約問題（以下、指定段落要約問題・一般化）では、具体例や数値など、具体的な記述が意味するところを要約に含める必要があるような段落を選定した。統合を割り当てた要約問題（以下、指定段落要約問題・統合）では、トピック・センテンス（文章の筆者自身による段落の要約）がない、あるいは明示的でない段落を選定した。Table 4.2に、冊子1で受検者に要約文の作成を求めた段落の例を示した。

文章全体要約問題では、指定段落要約問題では捉えきれない特徴を把握するため、設問の幹において特に指定を設けることはせず、「文章全体を要約してください」とのみ教示した。

さらに、解答に字数制限を設けると思考力や判断力を損なうとする実証的知見(安永他, 2013)に沿って、要約問題には字数制限を設けないこととした。指定段落要約問題では、要約文が1文程度で作成可能であることを考慮し「1文で解答してください」と教示した。一方で、文章全体要約問題では「複数文になってもかまいません」と教示した。加えて、本研究に参加する受検者の多くが日本語を母語とすることに配慮して、いずれの要約問題についても日本語で要約文を作成するよう、設問の幹において教示した。

#### 4.2.1.4 テスト冊子の構成

多枝選択式の英語文章読解問題1題(3項目)を第1問とし、記述式の英語文章要約問題3題のうち1題(4項目)を第2問とする大問2問構成のテスト冊子を計3種類作成した。第1問は3種類すべてのテスト冊子に収録する共通項目とし、第2問はテスト冊子によって異なる実験項目とした。

#### 4.2.1.5 手続き

解答データは、2016年10月～2017年4月の期間で収集された。受検者には、このテストが項目の性質を検討する目的で行われるものであること、受検者個人の評価とは全く関連しないものであることを伝えた。また、テストへの解答開始前において、解答データから個人が特定されないこと、研究参加をいつでも中断でき、そのことによる不利益はないことなどについて伝えたうえで、研究参加への同意を確認した。さらに、本研究の成

果を公表する際、個人を特定しない形で、受検者が作成した要約文を例として掲載することがあることについて伝えた上で、研究参加への同意を確認した。なお、本研究は名古屋大学大学院教育発達科学研究科の研究倫理審査委員会の承認を受けて実施された (ERB No. 16-827)

受検者には、テストへの解答を解答用紙にマーク・記入するよう教示した。テストへの解答時間は 35 分間としたが、時間内に解答が終了しない受検者がいた場合には、研究実施者が受検者の解答状況を見ながら、解答時間を 5 分程度延長することがあった。

#### 4.2.1.6 採点基準の作成・修正

受検者の要約文をもとに誤答の詳細を検討する前段階として、それぞれの要約問題における採点基準を作成した。採点基準の作成・修正ガイドラインとして、Haladyna & Rodriguez (2013) は、問題作成の段階で採点基準の作成を始めること、数名から実際に得られた解答に基づいて採点基準の適切さを評価することなどの重要性を指摘している。そこで本研究では、Haladyna & Rodriguez (2013) のガイドラインを参考に、少数の研究協力者から解答を得て暫定版の採点基準を作成した後に受検者に対してテストへの解答を求め、受検者の要約文をより適切に評価できるよう修正した上で、上記の研究協力者に最終版採点基準の確認を求める手続きをとった。以下では、それぞれの手続きの詳細を述べることとする。

なお、一般的に採点基準とは、受検者の解答に対して段階評定を行うための根拠資料のことを指すが、本研究では要約文における解答パターンを名義カテゴリとして考えていること、本研究の目的に照らした場合に詳細な段階評定値を与えることが必ずしも意味をもたないことなどの理由から、採点基準には段階評定の基準を含めず、正答の基準のみを示すこととした。ここで、正答の基準とは (1) それぞれの設問がターゲットとする認知過程、(2) ターゲットとした認知過程がその設問を解く際にはどのように反映しているか (設問のねらい)、(3) 解答例、(4) 要約文の正誤判断に必要な要素の 4 点のことを指す。研究協力者の事前の解答を踏まえて設定したのは、採点基準の (3) と (4) の部分であった。

まず、各項目の正答の基準を暫定的に決定するため、受検者の解答データを収集する前の段階に、英語を得意とする大学院生の研究協力者 6 名 (男性 2 名、女性 4 名) に要約問題 (第 2 問) のみの解答を求めた。研究協力者 1 名に対して 1 種類のテスト冊子を割り当てたため、テスト冊子 1 種類につき 2 名の研究協力者の解答データが得られた。研究協力者から得られた解答に基づいて、解答例の作成と正誤判断に必要な要素の特定を行い、暫定版採点基準を作成した。

次に、受検者 58 名分の解答データが得られた後、採点基準の修正を行った。具体的には、設問のねらいや必要な要素に関する記述のうち、受検者全体を通じてこれらの基準を満たした要約文が存在しなかったものについて、正答の要件から除外するなどの修正を

Table 4.2  
最終版採点基準の例（冊子 1・第 2 問）

設問	認知過程	設問のねらい	解答例	正誤判断に必要な要素
問 1	削除	休暇の計画を綿密に立てるときと、スケジュールをゆったりした計画にしたときとで、人々の休暇の楽しみ方が異なることを簡潔に表現できる	休暇の計画を綿密に立てた場合には、計画時のほうが楽しいが、計画に余裕を持たせた場合には、実際に休暇を過ごすときのほうが楽しい。	重要でない情報（膨大な研究データから得られた結果であることなど）が含まれていない
<p>There's a surprising amount of research on vacations and what aspects are satisfying. To begin with, vacation planning tends to bring happiness. Data show that people actually derive most happiness from a vacation in the planning phase. On its own, this finding would seem to recommend the creation of a complex schedule, rather than a spontaneous getaway. But if complex trips bring joy in the planning, they risk giving almost no happiness in the taking. <i>Harvard Business Review</i> reports that people get little to no happiness boost from vacations they consider stressful. Crowded schedules and tight connections may look exciting in the process of planning, but they suggest you return no happier than when you left. Only people with simple, relaxed trips seem to get a happiness boost. Think of this as the Vacation Paradox: You have to choose between being happy beforehand or happy afterward. Sorry.</p>				
問 2	統合	休暇をとることが賃金に及ぼす影響について、トピック・センテンスを作り出すことができる	有給休暇を消化しない人比べ、有給休暇を消化している人のほうが、昇給額やボーナスの額が多く、有給休暇をとることによるペナルティも受けていない。	有給休暇を使い切らない人と使い切る人の昇給やボーナスについて、対比的に述べられている
<p>Whether people take a vacation at all says a bit about career success, but the relationship is opposite from what one might imagine. According to a study from the U.S. Travel Association, American workers in 2013 gave up 169 million days of paid time off, worth \$52.4 billion. Was there a career benefit to giving up so much vacation time? There was not, according to the study. Those who left between 11 and 15 vacation day unused tended to receive smaller raises or bonuses than those who used all their vacation days. We don't know if this is because the vacationless employees were overstressed, or because incompetent employees who couldn't get their work done skipped vacation. But the finding does indicate that vacation takers are not paying a career penalty.</p>				
問 3	一般化	アメリカ人とヨーロッパ人の休暇に対する考え方を示した文、およびその周辺の文を、抽象的な表現に書き換えて要約することができる	仕事に価値を置くアメリカ人は有給休暇をあまり取らないが、休暇に価値を置くヨーロッパ人は有給休暇をたくさん取る。	アメリカ人・ヨーロッパ人の休暇に対する価値観、有給休暇の取り方について、具体的な記述（会話のはじめに相手によく尋ねることがら・有給休暇の日数に関するデータなど）を一般化した形で記述している
<p>Vacation also say a thing or two about one's country of origin. In America, the opening conversation topic with a stranger is frequently, "What do you do for a living?" For Europeans, the customary topic is "Where are you going on vacation this year?" Everyone has an answer, usually involving weeks and weeks at the beach or the mountains. The data provide support for the saying that Americans live to work, while Europeans work to live. The average American worker is offered just 16 days of paid holiday each year and has no legal guarantee to any time off. Compare this Italy (31 days), Spain (34 days) and Portugal (35 days).</p>				

行った。

採点基準の修正後、上記の研究協力者 6 名のうち、引き続き研究協力への同意が得られた 3 名(女性 3 名)に対し、修正後の採点基準の確認を求めた。研究協力者に対して本研究のデザインに関する説明を行ったうえで、3 種類すべてのテスト冊子を渡し、英語文章そのものと採点基準を突き合わせて両者が対応しているかどうか確認を行うよう依頼した。

以上の流れで作成・修正を行った採点基準を最終版の採点基準とし、次節に示す解答カテゴリのうち「正答」カテゴリの根拠を各項目ごとに示す資料として使用した。例として、Table 4.2 に冊子 1・第 2 問の最終版採点基準を示した。また、各設問の採点基準の下には、受検者に要約するよう求めた英語文章中の 1 段落を掲載した。

#### 4.2.1.7 解答カテゴリの生成

それぞれの設問に対する受検者の答案全体を確認し、受検者の要約文を評価するための解答カテゴリを作成した (Table 4.3)。

解答カテゴリの設定にあたり、以下の 2 点に留意した。第 1 に、正答以外のカテゴリについては、受検者から実際に得られた要約文と各認知過程の特徴をもとに、削除・一般化・統合それぞれの過程に特有の誤答を抽出した。このようなカテゴリ設定を行ったのは、後続する研究 7 において、削除・一般化・統合のそれぞれに見られる誤答の特徴を反映させた錯乱枝に関する検討を行うためである。

第 2 に、段落の要約文と文章全体の要約文の特徴を統一的な枠組みで検討するため、指定段落要約問題と文章全体要約問題の解答の両方を確認した上で、各認知過程の解答カテゴリを設定した。この手続きにより、各認知過程の解答カテゴリが指定段落要約問題・文章全体要約問題の両方で使用できるものとなり、文章全体要約問題独自のカテゴリを設定する必要が生じなくなる。本研究の要約問題に対する解答を共通の解答カテゴリで評価することになるため、指定段落要約問題で顕在化しやすい特徴、文章全体要約問題で顕在化しやすい特徴をそれぞれ明らかにすることが可能となる。

以下では、冊子 1・第 2 問に対する受検者の実際の要約文と各認知過程の特徴を対応づけながら、削除・一般化・統合それぞれに設定した解答カテゴリの詳細について述べる。

#### 削除に関する解答カテゴリ

文章要約モデルにおける「削除」のプロセスでは、重要でない情報や冗長な情報の取捨選択が想定されていることを踏まえ、正答以外に計 5 つの解答カテゴリ（重要でない情報の混入・必要な要素の不足・過度の削除・的外れ・読解の誤り）を設定した。

「重要でない情報の混入」は、段落や文章に含まれる情報を取捨選択する際に、重要でないものを取り除ききれなかった誤答として位置づけられる。例えば、「旅行の予定を詰めて立てると、計画の段階では楽しく、出発した後は疲れてストレスが溜まってしまうが、

Table 4.3  
解答カテゴリの一覧

認知過程	カテゴリ名	定義
削除	正答	正答の要件をすべて満たしており、重要でない情報・冗長な情報が取り除かれている
	重要でない情報の混入 必要な要素の不足	正答の要件はすべて含まれているが、重要でない情報も含まれている 正答の要件の一部のみが記述されているが、すべての要件を満たしているとは言えない
	過度の削除	要約文が過度に簡略化されており、正答の要件が満たされているかどうかの判断がつかない
	的外れ	正答の要件を全く満たしていない
	読解の誤り	正答の要件を満たしている程度にかかわらず、読解上の誤りを含む記述がある
一般化	正答	具体的な記述が上位の語に適切に言い換えられていたり、その記述の意味するところをまとめた表現がある
	具体的な記述の置き換え不足	具体的な記述（数値・固有名詞・並列されたことなどがらなど）が上位の語や表現に置き換えられていなかったり、具体的な記述の意味するところをまとめた表現がない
	不適切な一般化	具体的な記述を抽象的な語や表現に置き換えているが、その置き換えが適切でない
統合	正答	文章の筆者自身が段落の要約としてまとめている文を見つけ出して記述したり、文章の筆者の意図に沿ったわかりやすい要約文になっている
	一面的な記述	要約文として必要な要素の一部を取り出して記述しており、段落全体の内容を統合した要約文になっていない
	筆者の意図とのズレ	その段落での筆者の意図とは異なる方向の要約文になっている
	一貫性の欠如	要約文に一貫性がない

旅行の予定を簡単に立てておくと、計画の段階ではつまらないが旅行自体はリラックスした楽しいものになる」のような要約文は、採点基準 (Table 4.2) における設問のねらいを満たしてはいるものの、重要でない情報 (e.g., 実際の旅行では疲れてストレスが溜まってしまふ、旅行自体はリラックスしたものになる、など) も含んでいるため、この解答カテゴリにあてはまる。

「必要な要素の不足」は、段落や文章に含まれる重要な情報を重要でないと判断した誤答として位置づけられる。例えば、「休暇に関する調査において、人々は予定の詰まった計画を立てたとき、旅行に行ってからよりも計画を立てる段階でのほうが幸せを感じるということがわかっている」のような要約文は、旅行のスケジュールに余裕を持たせたときの人々の楽しみ方に関する記述がないため、設問のねらいを満たしていないと判断される。

「過度の削除」は、段落中の記述から読み取ることができた箇所のみをコンパクトに要約し、他の重要な箇所を書き落とした誤答として位置づけられる。例えば、「シンプルな休みこそ、幸せに過ごせる」のような要約文は、過度に簡潔化されているという点、旅行計画を綿密に立てるときの幸福感という、段落中の他の重要な箇所に言及されていない点

で、この解答カテゴリにあてはまる。

「的外れ」は、重要な情報を重要でないとする判断に加えて、重要でない情報を重要と判断した結果生じた誤答として位置づけられる。例えば、「休暇の計画を立てることは幸福なように思われるが、複雑な計画を立てるとかえって幸福から遠ざかる」のような要約文は、休暇の計画を綿密に立てた場合とゆったりとした計画にした場合で幸福感を感じるタイミングが異なるとする、当該段落の内容から逸脱していると判断される。なお、「的外れ」の特徴を含む要約文は他の過程を割り当てた要約問題でも確認されたが、情報の取捨選択という観点からこの特徴を捉えることが適切であると考え、削除過程に特有の誤答として位置づけた。

「読解の誤り」は、文章中に記述されている内容を解釈し間違えた誤答として位置づけられる。例えば、「最初に思っていた満足感を旅の後でも感じられるようにするのが大事だ」という要約文がこのカテゴリに該当する。ただし本研究では、この特徴を要約プロセスの誤答として捉えてはおらず、それ以前の文レベルの命題表象の構築の時点での誤答として捉えている。「読解の誤り」の特徴をもつ要約文は他の認知過程を割り当てた要約問題でも見られたが、文章要約モデルの中で削除過程が他の過程に先行していることを考慮し、削除過程に「読解の誤り」というカテゴリを含めた。

#### 一般化に関する解答カテゴリ

文章要約モデルにおける「一般化」のプロセスでは、並列されたことがらを上位概念の語に置き換えることが想定されていることを踏まえ、正答以外に2つの解答カテゴリ（具体的な記述の置き換え不足・不適切な一般化）を設定した。

「具体的な記述の置き換え不足」は、複写一削除方略 (Brown & Day, 1983) の使用により、段落や文章に含まれている具体的な記述をそのまま複写し、一般化した表現に置き換えなかった誤答として位置づけられる。例えば冊子1の第2問・問3では、アメリカ人・ヨーロッパ人の休暇に対する価値観や有給休暇の取り方について記述した段落の要約を求めた。ただし、アメリカ人・ヨーロッパ人の休暇に対する価値観に関する段落中の記述は、In America, the opening conversation topic with a stranger is frequently, “What do you do for a living?” For Europeans, the customary topic is “Where are you going on vacation this year?” (Table 4.2) のように極めて具体的である。受検者から得られた「アメリカでは見知らぬ人との最初の会話は『あなたは生活のために何をしているの?』であるのに対して、ヨーロッパ人は『今年の休暇はどこに行くの?』である」のような要約文は、複写一削除方略を使用し、まさに上の記述を抜き書きしたような解答となっていた。このような特徴をもつ要約文からは、アメリカ人・ヨーロッパ人の休暇に対する価値観を読み取ることができなかった様子がわかる。

「不適切な一般化」は、段落や文章中の具体的な記述を置き換えようとした表現が適切

でないという誤答として位置づけられる。例えば、「アメリカ人は働くために生き、ヨーロッパ人は生きるために働く」のような要約文には、段落中のやや一般化した表現に着目できているものの、休暇に対する価値観の違いを記述する文として、「アメリカ人は仕事に価値を置く」「ヨーロッパ人は休暇に価値を置く」のような表現は含まれていない。このような要約文は、具体的な記述を一般化する必要性については認識できているものの、より本質的な理解には至っていないことが読み取れる。

### 統合に関する解答カテゴリ

文章要約モデルにおける「統合」のプロセスでは、段落中で記述されている内容全体を代表するようなトピック・センテンスの産出が想定されていることを踏まえ、正答以外に計3つの解答カテゴリ（一面的な記述・筆者の意図とのズレ・一貫性の欠如）を設定した。

「一面的な記述」は、段落全体の命題の構成要素となる個々の命題群に適切な重みづけができなかった結果、一方に重きを置いて記述した誤答として位置づけられる。冊子1の第2問・問2では、決められた有給休暇をすべて消化できる人と消化しない人を比較した際の賃金の差や、有給休暇をとることによるペナルティの有無について言及されている段落を要約するよう求めた。この問題から得られた「長期休暇をたくさんとる人は、必ずしも、仕事で成功して収入の多い人だというわけではない」のような要約文は、有給休暇を消化する人のみに焦点を当てて記述しており、有給休暇を消化しない人との対比がなされていない点で、この解答カテゴリにあてはまる。

なお「一面的な記述」は、削除過程の「必要な要素の不足」と類似した部分も存在するが、本研究では、これら2つのカテゴリの背後に異なる認知過程を想定している。「必要な要素の不足」の背後には、要約に含めるべき情報を重要でないと判断した誤りが想定されるため、削除過程の誤答として位置づけられる。それに対し、「一面的な記述」の背後には、対比的な記述の片方に重きを置くような判断が想定され、トピック・センテンスとして段落の内容を代表しているとは見なせないという点において、統合過程の誤答として位置づけられる。このような観点から、対比的な記述の片方に重きを置くという「一面的な記述」の特徴を統合過程のカテゴリとして設定した。

次に、「筆者の意図とのズレ」は、段落中の記述に込められた筆者の意図を十分に反映していない誤答として位置づけられる。例えば、「仕事を休んでその分の給料をあきらめてまで、長期休暇をとることは割に合わないと考えて、休暇をとらない人が多くなっている」のような要約文は、有給休暇と仕事上の成功との関係に関する誤解を説明しようとした筆者の意図を十分に理解できたとはいえない。段落に込めた筆者の意図を十分に反映していない要約文は、段落を代表した文とは言えないという観点から、「筆者の意図とのズレ」を統合過程の誤りとして設定した。

「一貫性の欠如」は、その段落の内容を代表するトピック・センテンスとして一貫性に

欠けるという特徴をもった誤答として位置づけられる。例えば「研究によると、休暇を取った日数が比較的少ない人の方が受け取る給料が少ないとわかるが、休暇の少ない人は短い休みを嘆く仕事ができない人ばかりで、休暇中にも働くため、休暇そのものが人の実績に応じて罰則を定めているとはいえない」のような要約文は、必要な要素を含んでいるものの、段落中の記述をもとに構築されたミクロな命題群をより段落全体の命題として生成する過程で、ミクロな命題間の関係性に関する理解が十分でなかった結果、一貫性が低くなったものと考えられる。「一貫性の欠如」という特徴をもつ要約文も、他の過程を割り当てた要約問題で見られたが、情報の取舍選択や具体的な表現の一般化に関する誤りとして捉えることが難しく、段落全体の命題表象に一貫性がないという、統合過程に関連した誤りとして捉えることが可能であったため、統合過程の誤りとして「一貫性の欠如」を設定した。

#### 4.2.1.8 研究協力者による解答の分類

採点基準の確認を求めた研究協力者3名に対し、Table 4.3に示した解答カテゴリに沿って、それぞれの要約問題におけるすべての受検者の解答を分類するよう求めた。要約文に対する評定は、2017年6月～7月の期間に行われた。解答の分類の際、指定段落要約問題については、その問題がターゲットする認知過程の解答カテゴリの中からあてはまるカテゴリを1つ選ぶよう研究協力者に求めた。一方、文章全体要約問題については、削除・一般化・統合の解答カテゴリをすべて使用し、あてはまるカテゴリを1つ選ぶよう研究協力者に求めた。なお、研究協力者3名による解答の分類は独立に行われた。

#### 4.2.2 結果と考察

研究協力者から得られた要約文への評定データを基に、解答カテゴリの度数を要約問題の種類（指定段落要約問題・削除、一般化、統合、文章全体要約問題の4種類）ごとに算出し、受検者の要約文の特徴の出現頻度に関する把握を行った。なお本研究では、文章の違いを含みこんだ上で受検者の誤答の特徴を把握するため、異なるテスト冊子に収録された場合であっても、同じ種類の要約問題であれば評定データを結合して度数をカウントした。また、要約文が多様な特徴をもっている可能性を示唆する先行研究(邑本, 1992)があり、本研究においても要約文が複数の特徴をもつことが想定されるため、3名の研究協力者の評定データを結合して度数をカウントし、各カテゴリの出現頻度を把握した。

以下では、要約問題の種類ごとにそれぞれの解答カテゴリの度数を記した上で、先行研究との知見も踏まえながら、研究7で取り上げる誤答の特徴について述べることとする。なお、以下に示したそれぞれの度数表では、正答以外のカテゴリを度数の多かった順に並べて記載した。



**Table 4.4**  
指定段落要約問題・削除における各カテゴリの度数表

カテゴリ名	度数
正答	47
重要でない情報の混入	43
必要な要素の不足	35
読解の誤り	19
的外れ	18
過度の削除	12

#### 4.2.2.1 指定段落要約問題・削除

指定段落要約問題・削除における各カテゴリの度数表を Table 4.4 に示した。正答以外のカテゴリで顕著だったのは、重要でない情報の混入・必要な要素の不足であった。これら2つのカテゴリはともに、段落中の情報の取捨選択と密接な結びつきをもつ特徴であり、受検者の典型的な誤答であると考えられる。また、要約文に読解の誤りを含むケースも一定程度確認され、ミクロ構造として構築された命題が正しくないまま、情報の取捨選択に関する処理を行った受検者も存在した可能性が考えられる。要約文に含めた内容が的外れと評定されたケースも一定程度確認され、過度の削除がなされていると評定されたケースもわずかながら存在した。

韓国の大学生を対象とした Kim (2001) では、文章中で重要度の高いアイデア・ユニットが要約文中に多く含まれ、重要度の低いアイデア・ユニットが要約文中に占める割合が低かったことが報告されている。本研究において同様の分析を行ってはいないものの、日本の大学生を対象とした本研究で「重要でない情報の混入」・「必要な要素の不足」と評定されたケースが相対的に多かったことを考慮すると、要約文に含めるべき情報の取捨選択を適切に行うことができなかつた受検者が一定程度存在していたものと推察される。

#### 4.2.2.2 指定段落要約問題・一般化

指定段落要約問題・一般化における各カテゴリの度数表を Table 4.5 に示した。正答以外の2つのカテゴリはどちらも、一定程度受検者の要約文の中に見られたことがわかる。特に、不適切な一般化を含むと評定されたケースは正答と評定されたケースよりも多かった。

不適切な一般化に関しては、多くの受検者の要約文がこのような特徴をもっていると評定された可能性が考えられる。実際、冊子1の指定段落要約問題・一般化では「一般に、アメリカ人はヨーロッパ人と比較すると、娯楽より仕事を優先する傾向がある」という解

Table 4.5

指定段落要約問題・一般化における各カテゴリの度数表

カテゴリ名	度数
正答	59
不適切な一般化	67
具体的な記述の置き換え不足	39

Table 4.6

指定段落要約問題・統合における各カテゴリの度数表

カテゴリ名	度数
正答	44
一面的な記述	62
筆者の意図とのズレ	52
一貫性の欠如	5

答が得られた。この受検者の解答には、複製一削除方略の適用が見られず、段落中の内容の意味を抽象化して表現していることが見てとれる。ただし、アメリカ人とヨーロッパ人の休暇に対する「価値観」を述べた記述としては不十分であったため、「不適切な一般化」と評定されたものと考えられる。

一方、具体的な記述の置き換え不足に関して、Keck (2006) では、文章中に含まれる表現をそのまま複製したり、微修正を加えたりするだけの要約文がL2学習者に多く見られていた。本研究の受検者の中にも、エピソードや具体例を抽象化して理解することはせずに、段落中の表現の抜き書きで要約文を作成していた者がいた可能性が考えられる。

#### 4.2.2.3 指定段落要約問題・統合

指定段落要約問題・統合における各カテゴリの度数表をTable 4.6に示した。一面的な記述・筆者の意図とのズレと評定されたケースが多く、正答と評定されたケースと比べても多いという結果が得られた。一方で、要約文の内容の一貫性がないと評定されたケースは少なく、本研究の受検者がある程度一貫した要約文を作成していたことがうかがえる。

段落中の一部の記述に重きを置いている（一面的な記述）と評定された要約文の中には、二つのものを対比的に述べている箇所的一方のみが含まれ、その部分のみを取り上げて筆者の立場としているものが見られた。例えば、冊子2の指定段落要約問題・統合では、ボーナスを支給することによってモチベーションが上がる職業と、かえってモチベーションが下がる職業の特徴の違いについて述べられている段落の要約を求めた。この問題にお

いて、「従業員が自社や製品などに誇りを持っていなかったり、単純作業をしていたりする場合、ボーナスが有効な手段となる」のような要約文が得られた。この要約文の特徴では、筆者の意図を一部読み取れているが、ボーナスがかえってモチベーションを低下させる場合についての記述がないことから「一面的な記述」と評定されたものと考えられる。

筆者の意図とは異なる記述を含んでいる（筆者の意図とのズレ）と評定された要約文の中には、その段落中の記述から解釈できる範囲を超えて、恣意的に解釈しているものも見られた。例えば、上述の指定段落要約問題・統合（冊子2）では、「情熱がなければやる気が出ない一方、自分が考えた新しい企画ならあなたは従業員のやる気を出させる」のような要約文が得られた。実際、この要約文は、研究協力者から「筆者の意図とのズレ」と評定されたものであった。この段落で筆者は、金銭を与えることによりモチベーションが下がる場合と上がる場合について述べているにもかかわらず、上記の要約文では金銭に関する言及が見られないため、筆者の意図から逸脱していると判断される。

これらの特徴は、場合によっては削除過程と密接な結びつきをもつものであると考えられる。実際、一面的な記述という特徴は、必要な要素の不足という特徴の一部と考えることも可能であるし、筆者の意図とのズレという特徴は、読解の誤りという特徴の一部と考えることも可能である。研究7において、このような特徴を含んだ要約文を錯乱枝として提示する場合には、削除過程の影響も加味しながら、能力の高い受検者が「一面的な記述」や「筆者の意図とのズレ」を理由に誤答と判断できるような要約文を作成することの必要性が示唆された。

#### 4.2.2.4 文章全体要約問題

文章全体要約問題における各カテゴリの度数表を Table 4.7 に示した。正答以外のカテゴリのうち、最も度数が多かったのは必要な要素の不足であり、次いで過度の削除が多かった。どちらのカテゴリも、重要な情報を要約文に含めなかった誤りである点が共通していると言える。また、指定段落要約問題・削除とは大きく異なり、重要でない情報が混入していると評定されたケースはわずか1件のみであった。命題数が限られている段落の要約文を作成する場合には、重要でない情報を含めるという特徴が顕在的であったが、命題数が多い場合には、必要な要素を書き落とすという特徴が顕在的になる可能性が示唆される。

一般化過程・統合過程のカテゴリとして、具体的な記述の置き換え不足や一面的な記述・筆者の意図とのズレの特徴が要約文に含まれているとする評定が一定数確認された。ただし、評定全体から見ると度数は少なく、一般化過程や統合過程よりも削除過程の誤答に関する指摘が多く観測された。このような結果が得られた背景には、次の2つの可能性が考えられる。1つ目に、文章全体の要約では多くの命題を取り扱う必要があるため、情報の取捨選択に関する処理が優先され、一般化や統合の処理があまり求められていなかった

Table 4.7

文章全体要約問題における各カテゴリの度数表

カテゴリ名	度数
正答	20
必要な要素の不足 (削除)	46
過度の削除 (削除)	19
筆者の意図とのズレ (統合)	14
具体的な記述の置き換え不足 (一般化)	11
一面的な記述 (統合)	10
不適切な一般化 (一般化)	9
的外れ (削除)	2
読解の誤り (削除)	2
重要でない情報の混入 (削除)	1
一貫性の欠如 (統合)	1

たことが挙げられる。指定段落要約問題では、一般化や統合が特に重要となるような段落について要約文を作成するよう求めたため、一般化や統合に特有の誤答を抽出できたのに対し、文章全体要約問題では、一般化や統合が必要な段落を要約対象として含むものの、全体としては情報の取捨選択が優先されていた可能性が考えられる。

2つ目に、文章全体の要約に限らず、削除過程が一般化過程や統合過程の前処理として機能していたことが挙げられる。本章では、削除過程・一般化過程・統合過程を独立のプロセスとして扱っているが、実際には、削除過程を経た上での一般化過程・統合過程であることも想定されうる。本研究では、削除過程の解答カテゴリを一般化過程・統合過程の解答カテゴリと併用していたため、具体的な記述をより上位概念の語に置き換えて理解したり、筆者の意図に沿ったトピック・センテンスを産出したりする際の特徴が、削除過程における誤答に隠されてしまい、顕在化しなかった可能性も考えられる。また、研究協力者においても、削除過程の特徴により重きを置いて要約文の評定がなされ、他の過程の特徴が隠れてしまった可能性も考えうる。独立した認知処理として一般化過程・統合過程を扱うか、削除過程を含めた一般化過程・統合過程を扱うかはテストの目的や仕様に依拠するものであるが、本研究のように各過程を独立に扱う立場に立つならば、この点を十分に考慮する必要があるものと考えられる。用いる文章によっては、一般化や統合が必要な段落を含みながらも、指定段落要約問題・削除に類似した問題となる可能性を考慮する必要がある。

#### 4.2.2.5 本研究から得られた示唆と研究7で取り上げる錯乱枝

研究7では、本研究で明らかにされた誤答の特徴を踏まえて、英語文章の要点把握能力を測る多枝選択式項目の錯乱枝を作成する。項目作成の詳細は4.3.1.3節で述べるが、多枝選択式項目の選択枝数は3枝（正答選択枝1つ・錯乱枝2つ）で十分であるとする知見(Haladyna & Rodriguez, 2013)に基づいて、各認知過程の解答カテゴリから、錯乱枝に含める誤答の特徴2種類を以下の観点で抽出した。

削除過程では、度数が多かった「必要な要素の不足」と「重要でない情報の混入」を取り上げることとした。他のカテゴリを取り上げない理由としては、以下の3点が挙げられる。第1に、「過度の削除」の特徴をもつ要約文は、他の特徴をもつ要約文と比べて短い傾向にあると考えられる。Haladyna & Rodriguez (2013)では、選択枝の長さの違いが正答を選ぶためのヒントとなる可能性があることを示唆している。本研究ではこの点を考慮し、「過度の削除」を研究7の検討対象から除くこととした。第2に、「的外れ」については、このような特徴をもつ要約文を意図的に作成することが難しいことから、研究7では取り上げなかった。第3に、「読解の誤り」は上述のように、本研究では要点把握に至る以前の誤りとして捉えているため、要点把握能力を反映した誤答として不適切であると判断した。

一般化過程では、もともと誤答の特徴として2種類の解答カテゴリ（具体的な記述の置き換え不足・不適切な一般化）のみの設定であったため、これら2つの特徴をそのまま取り上げることとした。

統合過程では、「一面的な記述」と「筆者の意図とのズレ」を取り上げることとした。「一貫性の欠如」を取り上げなかった理由としては、このような特徴をもつ要約文を読むことが不必要な認知負荷を要求する可能性があった点が挙げられる。

### 4.3 研究7：英語文章の要点を把握する問題における錯乱枝の検討

研究7では、本章の第2の課題である、受検者の典型的な誤答の特徴を反映した錯乱枝の選択率に関する検討を行う。この研究課題の遂行にあたっては、研究6で明らかにされた誤答の特徴を踏まえることとする。

#### 4.3.1 方法

##### 4.3.1.1 受検者

愛知県の大学に通う学生を受検者とした。研究協力に同意のなかった者、すべての項目に解答のなかった者を取り除いた上で、414名(男性68名、女性346名)のデータを分析対象とした。受検者は全員、英語を第二言語とする者であった。

Table 4.8  
研究 7 で用いた文章一覧

冊子	テーマ	リーダビリティ	語数
冊子 1	ホームシック	9.48	742
冊子 2	言語と認識	13.83	508
冊子 3	人間の行動の変容	8.42	755
冊子 4	オランダとチーズ	8.10	920
冊子 5	デジタル時代の人間の心	12.57	876
冊子 6	気候変動と動物の絶滅	11.84	567
共通	絶対菜食主義	8.64	643

#### 4.3.1.2 材 料

研究 6 と同様の基準で、大学の入学試験問題の英語文章読解問題を 7 題選定した。7 題の英語文章読解問題はすべて、英語文章の要点を把握する能力を測定する問題として、本研究の目的に沿って作成した。研究 7 で用いた英語文章の一覧を Table 4.8 に示した。

#### 4.3.1.3 テスト項目の作成

本研究では、英語文章 1 題につき 3 項目の多枝選択式問題を作成した。研究 6 の指定段落要約問題と同様、削除・一般化・統合の各プロセスを項目と 1 対 1 対応させ、削除のプロセスに焦点を当てた問題（以下、削除問題）1 項目、一般化のプロセスに焦点を当てた問題（以下、一般化問題）1 項目、統合のプロセスに焦点を当てた問題（以下、統合問題）1 項目を作成した。これら 3 種類の項目を英語文章 6 題分、計 18 項目を作成した。

すべての項目において、正答選択枝 1 つ、誤答選択枝 2 つの 3 枝選択問題とした。誤答選択枝に反映させる誤答として、研究 6 で得られた各認知過程の誤答の特徴のうち出現頻度の高かったものを 2 つずつ抽出した。削除問題の誤答選択枝には「重要でない情報の混入」・「必要な要素の不足」を、一般化問題の誤答選択枝には「具体的な記述の置き換え不足」・「不適切な一般化」を、統合問題の誤答選択枝には「一面的な記述」・「筆者の意図とのズレ」をそれぞれ反映させた。

削除問題・一般化問題・統合問題の作成にあたり、設問や選択枝の作成基準を明確にするため、研究 6 における指定段落要約問題の作成手続きに類似した手順をとった。具体的には、Table 4.9 に示した表を作成しながら、要点把握を求める段落の選定や要点の把握に必要な要素の確定、選択枝に含めるべき誤答についての検討を行った。項目作成手続きの大まかな流れとしては、1) 要点の把握を求める段落の選定、2) その段落を要約する際に必要な要素の抽出、3) 正答選択枝の作成、4) 錯乱枝の作成と誤答の理由の明確化であった。

第 1 に、削除・一般化・統合のそれぞれのプロセスを各項目に割り当て、研究 6 と同様

Table 4.9  
項目の仕様を記した一覧表の例 (冊子 1・第 2 問)

項目	段落	認知過程	必要な要素	選択枝種別	選択枝本体	誤答の理由
問 1	2	削除	(1) 経済的な困難や必要に迫られ、家族のもとを離れて働く人々がいる (2) (近年の交通網の発達で) 遠距離への移動が容易になった (3) 家族のもとを離れて働く人々の多くは、ホームシックになっている	正答	As the global mobility become available, more and more people who need to earn money travel abroad easily, but some of them feel lonely living apart from their family members.	(1) ~ (3) を満たしている
				混入	According to the survey on emotional experiences of immigrants for about ten years, most of them who met their necessity to succeed in another country thanks to the global mobility suffer from homesickness.	重要でない情報が混入している
問 2	4	統合	(1) 20 世紀初頭に通信技術が発達し、遠くにいる家族と頻繁に連絡を取り合えるようになった (2) ホームシックは、通信技術の発達後も移住者を悩ませ続けている	不足	Due to the easiness of global transportation, people who are poor and distant from members of their family to find more profitable work in another country tend to get opportunity and earn much money.	(3) の内容が含まれていない
				正答	Homesickness has existed in immigrants even if information technology advanced in the beginning of the 20th century and a regular contact with their family member became possible.	(2) の内容が含まれていない
問 3	7	一般化	具体例：アメリカに移住したカルメンは、安価なインターネット電話で頻繁に家族と電話していたため、常に家族の動向を意識するようになり、寂しさが増した (1) 通信技術の発達で、家族といつでも連絡がとれるようになった (2) 家族と頻繁に連絡を取り合えるようになると、家族と離れていることを常に意識するようになる (3) 家族のことが忘れられないと、ホームシックになる	ズレ	Owing to the advance in technology early in the 20th century, people can communicate with their family constantly and homesickness seemed to be less common, but nowadays immigrants tend to suffer from it again.	筆者が <b>continue</b> としているところを <b>again</b> としている
				正答	Modern technology that enables us to contact regularly with distant people remind immigrants of their family, and stronger feelings of displacement occur to them.	具体例をそのまま記述している
				不適切	Recent information technology made immigrants to the United States get homesick because a regular phone call with lower price feel that they are separated from members of their family.	アメリカに移住する人限定の話題になっている

に、割り当てた認知プロセスが要点把握の際に重要となるような段落の選定を行った。削除問題では、類似する内容が複数回述べられていたり、周辺的な内容を多く含んでいたりするなどの特徴をもつ段落を選定した。一般化問題では、具体例や数値など、具体的な記述の意味を上位概念の語に置き換えて理解することが求められる段落を選定した。統合問題では、トピック・センテンスが書かれていない、あるいは明示的でない段落を選定した。

第2に、各項目に割り当てた認知過程の特徴を念頭に置きながら、段落の要点を把握する際に必要な要素の同定を行った。削除問題では、段落中に含まれる記述内容を整理した上で、必要な要素のみを抽出した。例えば、Table 4.9 に示した冊子1・第2問・問1は削除問題として作成されたものであり、段落の中には以下の5つの内容が含まれていた。

1. 貧困やその他の必要性から故郷を離れようと思う心が生まれ、国をまたぐ移動が可能になったことにより、その欲求はさらに高まっている
2. 一般には、人は世界中のどこよりも、自分の国にいたほうがよいと考えている
3. 仕事などのために故郷を離れるという考え方はいまや、グローバル経済の中核をなすものとして受け入れられている
4. 故郷を離れることで機会に恵まれたり、収入をもたらしたりするが、一方で心理的なコストも高い
5. 移民の感情や体験について10年近くにわたり調査した研究では、よりよい職業を求めて故郷を離れた人の多くが、離郷による寂しさを感じている

冊子1・第2問で提示した文章では、全体にわたってホームシックについて述べられており、特にこの段落の中核は、故郷を離れた人々が故郷への思いを募らせ、寂しさを感じているという点である。上記の内容の中で、2番目や3番目に示した内容は重要でない情報であり、1番目・4番目・5番目の内容のみが必要な要素ということになる。削除問題では、上記のような手続きを経て必要な要素の取捨選択を行った。

一般化問題では、段落中の具体的な記述とそれに対応する一般化表現との対応を明確にしながら、必要な要素の特定を行った。例えば、冊子1・第2問・問3では、「新しいテクノロジーが離郷による寂しさを強める (these new technologies actually make feelings of displacement stronger)」という文から始まっており、その次に Maria Elena Rivera という人物のエピソードが挿入されている。このエピソードは段落中で、新しいテクノロジーが寂しさを強めてしまう理由を具体的に説明する役割を果たしている。このような構造をもつ段落の要点把握にあたっては、エピソードに含まれている事例を抽象化して理解する必要が生じる。エピソードを踏まえると、新しい通信技術の発達により、家族といつでも連絡をとれるようになったこと、そのことにより家族の動向が手に取るようになるため、自分がそこにはいないことを常に意識してしまうことなどが読み取れる。そこで、冊子1・第2問・問3では、必要な要素のひとつとして、エピソードから導出される「家族と頻りに



連絡を取り合えるようになる」と、家族と離れていることを常に意識するようになる」点を含めた。また、その前提にある「新しい通信技術により、家族といつでも連絡がとれるようになった」点や、その結果としての「家族のことが忘れられずホームシックになる」点も必要な要素として含めることとした。

統合問題では、段落中の記述全体を反映させるようにして、必要な要素を定めた。例えば、冊子1・第2問・問2では、情報技術の発展がホームシックに与えてきた影響を歴史的な観点から述べた段落の要点を理解するよう求めた。この段落では、アメリカに移住してきた多くの人々がホームシックになった19世紀の状況が、20世紀における情報技術の発展により解決されたように思われているが、実態としてはいまもなお多くの人々がホームシックになっていると記述されている。このような整理を踏まえ、(1) 20世紀に情報技術の発展があった点、(2) 情報技術の発展後もホームシックは移住者を悩ませ続けている点を必要な要素として設定した。

第3に、各項目の必要な要素をすべて反映させるようにして、正答選択枝を作成した。

最後に、各認知過程における典型的な誤答の特徴を反映させた錯乱枝を作成した。削除問題では、「必要な要素の不足」錯乱枝として、必要な要素の一部を書き落とした要約文を提示し、「重要でない情報の混入」錯乱枝として、すべての必要な要素に加えて周辺的な情報を追加した要約文を作成した。一般化問題では、「具体的な記述の置き換え不足」錯乱枝として、英語文章中の記述を上位概念の語に置き換えていない表現を含んだ要約文を提示し、「不適切な一般化」錯乱枝として、置き換え後の上位概念の語が適切でない要約文を作成した。統合問題では、「一面的な記述」錯乱枝として、段落中の記述の一部に過度な重きを置いた要約文を提示し、「筆者の意図とのズレ」錯乱枝として、段落中の筆者の意図から逸脱した要約文を作成した。

#### 4.3.1.4 手続き

解答データは、2017年10月～2018年1月の期間で収集された。受検者には、このテストが受検者自身を評価する目的で行われるものではなく、テスト項目の性質について検討する目的で行われるものであることを伝えた。また、受検者に対する倫理的配慮として、テスト冊子が研究室内で厳重に保管されること、研究参加を中断しても受検者自身が不利益を被ることがないことなどを伝えた後、受検者の同意を確認した。なお、本研究は名古屋大学大学院教育発達科学科の研究倫理審査委員会の承認を受けて実施された(ERB No. 17-1039)。

テストへの解答にはマークシートを使用することとし、正しいと思う選択枝の丸番号をマークシート上で塗りつぶすよう受検者に教示した。解答時間は30分間であった。

Table 4.10  
項目分析の結果（研究7）

第1問	項目	選択枝	選択率	点双列相関		項目	選択枝	選択率	点双列相関	
共通項目	問1	1 (正 答)	.263	.399						
		2 (不適切)	.222	-.197						
		3 (具体的)	.514	-.187						
	問2	1 (混 入)	.280	-.191						
		2 (正 答)	.370	.418						
		3 (不 足)	.350	-.242						
	問3	1 (一面的)	.295	-.191						
		2 (ズ レ)	.280	-.252						
		3 (正 答)	.425	.405						
第2問	項目	選択枝	選択率	点双列相関		項目	選択枝	選択率	点双列相関	
冊子1 (72名)	問1	1 (正 答)	.333	.280		冊子4 (69名)	1 (不 足)	.362	-.411	
		2 (混 入)	.375	-.179			問1	2 (正 答)	.362	.621
		3 (不 足)	.292	-.100			3 (混 入)	.275	-.226	
	問2	1 (一面的)	.347	-.327			問2	1 (一面的)	.464	-.140
		2 (ズ レ)	.333	-.166			2 (ズ レ)	.217	-.184	
		3 (正 答)	.319	.502			3 (正 答)	.319	.313	
	問3	1 (不適切)	.264	-.231			問3	1 (正 答)	.246	.425
		2 (正 答)	.444	.421			2 (不適切)	.304	-.144	
		3 (具体的)	.292	-.236			3 (具体的)	.449	-.235	
冊子2 (68名)	問1	1 (一面的)	.500	-.189		冊子5 (68名)	1 (混 入)	.265	-.241	
		2 (正 答)	.273	.261			問1	2 (不 足)	.412	-.243
		3 (ズ レ)	.227	-.052			3 (正 答)	.324	.483	
	問2	1 (具体的)	.455	-.310			問2	1 (正 答)	.338	.653
		2 (不適切)	.258	-.006			2 (不適切)	.382	-.234	
		3 (正 答)	.288	.347			3 (具体的)	.279	-.435	
	問3	1 (正 答)	.415	.518			問3	1 (ズ レ)	.269	-.455
		2 (不 足)	.262	-.368			2 (一面的)	.224	-.092	
		3 (混 入)	.323	-.200			3 (正 答)	.507	.480	
冊子3 (68名)	問1	1 (混 入)	.284	-.121		冊子6 (69名)	1 (不 足)	.435	-.390	
		2 (正 答)	.463	.532			問1	2 (混 入)	.246	-.013
		3 (不 足)	.254	-.484			3 (正 答)	.319	.427	
	問2	1 (不適切)	.212	-.250			問2	1 (具体的)	.203	-.269
		2 (具体的)	.379	-.262			2 (正 答)	.420	.568	
		3 (正 答)	.409	.466			3 (不適切)	.377	-.355	
	問3	1 (ズ レ)	.364	-.281			問3	1 (正 答)	.348	.561
		2 (正 答)	.379	.284			2 (一面的)	.319	-.445	
		3 (一面的)	.258	-.005			3 (ズ レ)	.333	-.127	

## 4.3.2 結 果

### 4.3.2.1 項目分析

多枝選択式テストを用いたこれまでの研究と同様、受検者の応答データを2値採点した上で、項目分析を行った。ただし、これまでの分析とは異なり、第1問（共通項目）も英語文章の要点把握問題として本研究が独自に作成したものであるため、第1問・第2問と

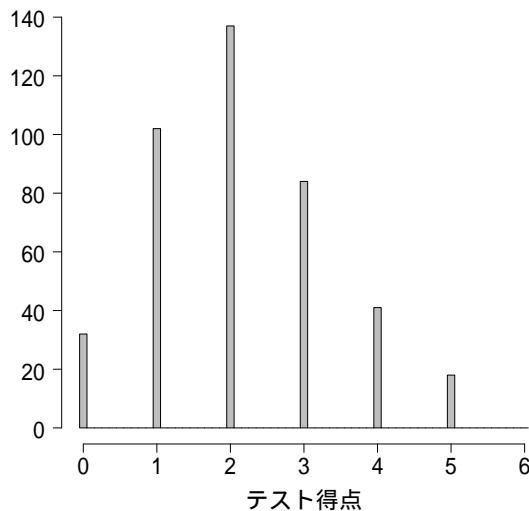


Figure 4.1. 受検者の得点分布（研究 7）

Table 4.11. 冊子ごとの得点の記述統計量

冊子	<i>M</i>	<i>SD</i>
冊子 1	2.222	1.129
冊子 2	2.000	1.106
冊子 3	2.294	1.259
冊子 4	2.029	1.236
冊子 5	2.265	1.389
冊子 6	1.971	1.294

もに選択枝単位での分析を行った。分析にあたり，困難度の指標として各選択枝の選択率を，識別力の指標としてその選択枝への選択状況（選択=1，非選択=0）と潜在特性値の EAP 推定値との点双列相関係数を算出した。

項目分析の結果を Table 4.10 に示した。全体として，正答選択枝の選択率を見ると，極端な値をとっているものは存在せず，点双列相関係数も十分な値を示していた。誤答選択枝に関して，選択率・点双列相関係数ともに十分な値を示しているものが多く見られたが，一部の選択枝において点双列相関係数が 0 に近い値を示していた（冊子 3・問 3・選択枝 3，冊子 5・問 3・選択枝 2，冊子 6・問 1・選択枝 2）。このうち，冊子 3・冊子 5 において誤答選択枝の点双列相関が低かった選択枝は，いずれも統合のプロセスを割り当てたテスト項目であり，一面的な記述錯乱枝であった。冊子 6 において点双列相関の低かった選択枝は，重要でない情報の混入錯乱枝であった。

#### 4.3.2.2 受検者の得点分布

受検者のテスト得点のヒストグラムを Figure 4.1 に示した。受検者の得点の平均は 2.130 点，標準偏差は 1.238 点であった。

テスト冊子ごとにみた受検者の得点について検討を行うため，テスト冊子ごとに得点の平均値と標準偏差を算出した (Table 4.11)。受検者の得点の平均値をテスト冊子ごとにみると，2.00 点～2.30 点の間に収まっており，テスト冊子間で得点の平均値に大きな差があるとは見なせなかった。

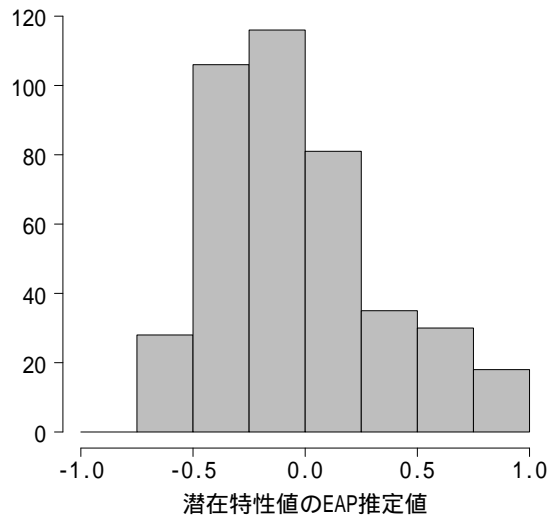


Figure 4.2. 受検者の潜在特性値の EAP 推定値のヒストグラム (研究 7)

#### 4.3.2.3 受検者の潜在特性値の推定

多枝選択式テストを用いたこれまでの研究と同様、共通項目（第 1 問）をアンカーとし、項目応答理論における 1PLM を用いて、受検者の潜在特性値を EAP 推定した。推定にあたって、受検者母集団の潜在特性値  $\theta$  の事前分布は標準正規分布とした。なお、これまでの研究と同様、本研究においても受検者によって割り当てられたテスト冊子が異なるため、割り当てられなかったテスト冊子に含まれる項目への応答を欠測扱いとしたデータ行列を推定に使用し、同時尺度調整法を適用して潜在特性尺度の調整を行った。

受検者の潜在特性値の EAP 推定値のヒストグラムを Figure 4.2 に示した。受検者の潜在特性値の平均は 0.00、標準偏差は 0.34 であった。

#### 4.3.2.4 受検者の能力群の構成

多枝選択式テストを用いたこれまでの研究と同様、受検者の潜在特性値の EAP 推定値に基づいて能力群を構成した。能力群の分割には、Kelley (1939) の方法を適用し、受検者の潜在特性値の推定値の分布の 27 パーセンタイル点 (-.278)、73 パーセンタイル点 (.238) を求めた上で、27 パーセンタイル点以下の受検者 119 名を能力低群 ( $M = -0.393$ ,  $SD = 0.125$ )、27 パーセンタイル点から 73 パーセンタイル点までの受検者 187 名を能力中群 ( $M = -0.009$ ,  $SD = 0.128$ )、73 パーセンタイル点以上の受検者 108 名を能力高群 ( $M = 0.449$ ,  $SD = 0.200$ ) として、受検者を 3 群に分割した。

#### 4.3.2.5 削除問題・一般化問題・統合問題における錯乱枝の選択率の検討

これまでの章と同様、それぞれの錯乱枝の選択率に関する検討にあたっては、英語文章の違いに依拠しない知見を提供するため、異なる文章を提示した場合であっても、同一の認知過程が割り当てられた項目であれば解答データを結合し、分析を行った。ただし、本研究では第1問（共通項目）・第2問（実験項目）の両方に削除問題・一般化問題・統合問題が存在するため、同一の認知過程を割り当てた問題を結合した場合、同一の受検者から得られた2種類の応答を1つの分析で扱うこととなる。このデータを対応のあるデータと見なして分析する方法もあるが、本論文では一貫して実験項目のみを分析対象としていた点を考慮し、本研究では共通項目として位置づけた第1問の削除問題・一般化問題・統合問題を分析には含めず、実験項目として位置づけた第2問の削除問題・一般化問題・統合問題のみを分析対象とした。Figure 4.1に、各認知過程を割り当てた問題別のトレースラインを示した。

以下では、トレースラインの特徴の把握とオッズ比の検討を行うこととする。なお、オッズ比の検討にあたってはこれまでの章と同様、正答選択枝の選択率に対するそれぞれの錯乱枝の選択率のオッズを能力群ごとに構成し、そのオッズを能力群間で比較するようなオッズ比を構成した。また、95%信頼区間については、対数オッズ比の標本分布が漸近的に正規分布に従う性質を利用した。

本論文における分析番号の統一性の観点から、本研究の分析を分析7とし、削除問題の結果を分析7-1に、一般化問題の結果を分析7-2に、統合問題の結果を分析7-3に示した。

#### 各要点把握問題のトレースライン

削除問題では、必要な要素の不足錯乱枝・重要でない情報の混入錯乱枝ともに、能力水準が高い場合に選択率が低い傾向が見られた。特に、必要な要素の不足錯乱枝の選択率のほう折れ線の傾きが大きくなっているように見受けられる。一般化問題では、具体的な記述の置き換え不足錯乱枝・不適切な一般化錯乱枝ともに類似した推移が見られた。2種類の錯乱枝ともに、能力低群・中群において選択率が高く、能力高群において選択率が低い傾向が見てとれる。統合問題では、2種類の錯乱枝の推移が類似しており、どちらの錯乱枝の選択率も能力低群では選択率が同程度に高いが、能力中群・高群では一面的な記述錯乱枝と筆者の意図とのズレ錯乱枝の選択率の差が拡大していることがうかがえた。

#### 分析7-1：削除問題（第2問）の錯乱枝の選択率に関する検討

分析7-1におけるオッズ比の推定結果をTable 4.12に示した。必要な要素の不足錯乱枝に関して、能力中群—高群間、能力低群—中群間ともにオッズ比の値が大きくなっていた。この結果から、必要な要素の不足錯乱枝の選択率が能力群間で大きく異なることが示され

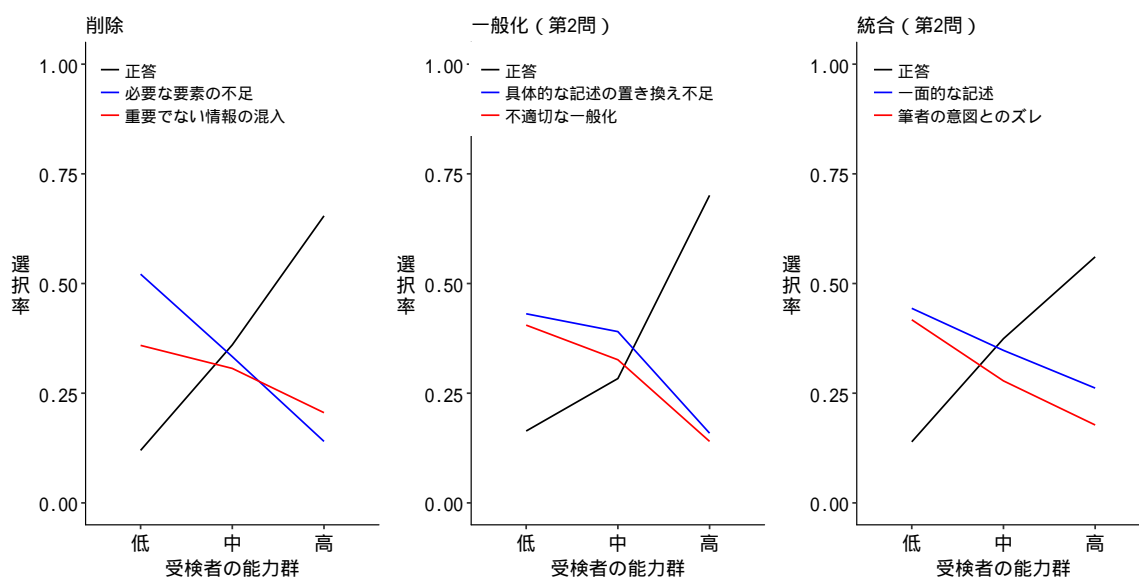


Figure 4.3. 研究 7 におけるトレースライン (第 2 問)

Table 4.12

分析 7-1 におけるオッズ比の推定結果

比較対	必要な要素の不足錯乱枝		重要でない情報の混入錯乱枝	
	OR	95%CI	OR	95%CI
能力中群—能力高群	4.318	[2.241, 8.321]	2.707	[1.493, 4.908]
能力低群—能力中群	4.709	[2.396, 9.255]	3.526	[1.750, 7.104]

た。重要でない情報の混入錯乱枝に関しても、能力低群—中群間で大きなオッズ比が確認された。重要でない情報の混入錯乱枝は、能力中群よりも能力低群の選択率のほうが高くなることが示唆された。能力中群—高群間においても十分大きなオッズ比が得られたが、能力低群—中群間のオッズ比と比べると小さな値であった。

全体として、削除過程の典型的な誤答を反映させた 2 種類の錯乱枝はいずれも、能力群間で選択率の差が見られた。なかでも、必要な情報の不足錯乱枝は能力群間での選択率の差が大きく、受検者の能力を識別するのに有効な錯乱枝であることが示唆された。

#### 分析 7-2：一般化問題（第 2 問）の錯乱枝の選択率に関する検討

分析 7-2 におけるオッズ比の推定結果を Table 4.13 に示した。トレースライン (Figure 4.3) でも示されていたように、具体的な記述の置き換え不足錯乱枝・不適切な一般化錯乱枝ともに、能力中群—高群間でオッズ比が大きくなっており、能力低群—中群では相対的にオッズ比の値が小さいことが示された。

#### 分析 7-3：統合問題（第 2 問）の錯乱枝の選択率に関する検討

Table 4.13

分析 7-2 におけるオッズ比の推定結果

比較対	具体的な記述の置き換え不足錯乱枝		不適切な一般化錯乱枝	
	OR	95%CI	OR	95%CI
能力中群—能力高群	6.077	[3.223, 11.458]	5.755	[2.958, 11.195]
能力低群—能力中群	1.910	[1.012, 3.608]	2.149	[1.125, 4.107]

Table 4.14

分析 7-3 におけるオッズ比の推定結果

比較対	一面的な記述錯乱枝		筆者の意図とのズレ錯乱枝	
	OR	95%CI	OR	95%CI
能力中群—能力高群	1.990	[1.135, 3.488]	2.346	[1.251, 4.398]
能力低群—能力中群	3.433	[1.783, 6.610]	4.038	[2.067, 7.892]

分析 7-3 におけるオッズ比の推定結果を Table 4.14 に示した。一面的な記述錯乱枝・筆者の意図とのズレ錯乱枝の両者とも、能力低群—中群間において大きな値のオッズ比が確認され、能力中群—高群間のオッズ比の 2 倍程度の大きさであった。トレースラインにおいて、2 種類の錯乱枝が類似した推移をたどっていたことに対応し、オッズ比の大きさも類似していた。この結果から、一面的な記述錯乱枝・筆者の意図とのズレ錯乱枝ともに、能力低群—中群間の識別に寄与しうることが示された。

いずれの能力群間においても、2 種類の錯乱枝ともに選択率の差が見られ、特に能力低群—中群間で選択率の大きな差が認められた。一方で、2 種類の錯乱枝でオッズ比の大きさのパターンが類似しており、錯乱枝の種類によらず類似した選択率の推移を示していた。

### 4.3.3 考 察

研究 7 の目的は、研究 6 で明らかにされた誤答の特徴を反映させた錯乱枝を作成し、各錯乱枝の選択率を能力群別に検討することであった。研究 6 の指定段落要約問題と同様、Kintsch & van Dijk (1978) の文章要約モデルにおける削除・一般化・統合のうちいずれかをテスト項目に割り当て、そのプロセスに付随する典型的な誤答の特徴 2 つを取り上げた上で、それぞれの特徴を反映させた錯乱枝を作成し、選択率に関する検討を行った。以下では、認知過程ごとに結果を整理しながら考察を進めることとする。

#### 4.3.3.1 削除問題

分析 7-1 の結果、削除問題においては、必要な要素の不足錯乱枝の選択率が能力群間で大きく異なるという結果が得られた。研究 6 では、同様の特徴をもつ要約文が多くみられ

たが、多枝選択式項目を用いた本研究では、このような特徴をもつ要約文を錯乱枝として提示した場合に、能力群間での選択率の差が大きいという結果が得られた。特に、能力低群と能力中群の間で選択率が大きく異なっていることが示された。また、重要でない情報の混入錯乱枝についても、能力群間で十分に大きなオッズ比が得られ、必要な要素の不足錯乱枝ほどではないものの、能力群間で選択率が異なることが示された。

必要な要素の不足錯乱枝に関して、能力の低い受検者は、段落中に含まれる情報の重要度に対する判断が十分でなかったために、必要な要素が要約文中に含まれていなかったとしても正しいと判断していた可能性が考えられる。一方、能力の高い受検者は、段落中の情報の重要度を適切に判断できるため、必要な要素を欠いた要約文を不適切であると判断することは容易であった可能性がある。このように、能力の低い受検者と高い受検者との間で、必要な要素の不足の検出に違いが見られたために、必要な要素の不足錯乱枝が能力群間で選択率の大きな差をもたらしていた可能性が考えられる。この結果は、Kintsch & van Dijk (1978) や Brown & Day (1983) のモデルと整合するものであった。

それに対して、重要でない情報の混入錯乱枝は、必要な要素をすべて満たすように作成されたものであるため、能力の高い受検者においても「必要な要素は含まれている」として重要でない情報の混入を許容した者が一定程度存在していた可能性も考えられる。削除過程では字義通り「削除」が求められ、要点把握の対象となっている段落の命題表象を構築する際、重要でない情報を取り除くという処理が必要となる。本研究では、重要でない情報の混入の錯乱枝に関して、能力中群—高群間の選択率の差は相対的に小さいという結果が得られ、重要でない情報を取り除くという処理に関して、能力中群と高群との間では特筆すべき大きな違いがなかった可能性が考えられる。

本研究の削除問題では、段落の要点を理解するにあたって必要な要素とそうでない要素を区別し、適切な要約文を選択することが求められる。要約文を作成する際には、情報の重要度に関する判断が情報の取捨選択の前処理として重要であるとされており (Hidi & Anderson, 1986)、本研究の結果を踏まえると、要点把握においても重要であると考えられる。特に、必要な要素の一部を欠いた要約文を錯乱枝として提示することにより、要点把握の中でも必要な要素の特定のプロセスをうまく抽出することができる可能性が示されたと言える。

#### 4.3.3.2 一般化問題

分析 7-2 の結果、一般化問題においては、具体的な記述の置き換え不足錯乱枝・不適切な一般化錯乱枝ともに、能力群間での選択率の差が見られた。2種類の錯乱枝とも選択状況が類似しており、能力低群・中群で選択率が高く、能力高群で選択率が大きく下がる傾向が確認された。

このような結果が得られた背景には、一定程度高い能力水準に達しないと、具体的な表



現を適切に抽象化して要点を把握することが困難であった可能性が考えられる。本研究の一般化問題では、具体例が多く含まれる段落を要点把握の対象として指定しており、具体例の意味するところをそのまま記した要約文ではなく、文章全体のトピックに合うよう、上位概念の表現に置き換えた要約文が正答となる。このような中で、能力低群・中群のうち、具体的な記述の置き換え不足錯乱枝を選択した受検者は、具体例の抽象化を通して要点を把握することの必要性を認識していなかったため、段落中の具体例を置き換えた記述がなく、段落中の一部を抜き書きしたのみの錯乱枝を選んでいたことが考えられる。一方で、能力低群・中群の中には、不適切な一般化錯乱枝を選択した受検者も一定程度存在した。このような受検者は、具体例をいくらか言い換える必要性については認識していたことが考えられるが、適切な一般化表現を含む要約文を選ぶことができなかったものと推察される。このように、能力低群・中群の受検者の中には、具体例の抽象化の必要性を認識していない者と、その必要性を認識しつつも適切な一般化表現についての判断を誤った者が混在していることが示唆される。能力高群では、能力中群と比べてもこれらの錯乱枝の選択率が低いという結果が得られ、具体的な表現を抽象化することの必要性を認識した上で、一般化表現の適切さに関する判断も行うことができたことが考えられる。

#### 4.3.3.3 統合問題

分析 7-3 の結果、統合問題においては、一面的な記述錯乱枝・筆者の意図とのズレ錯乱枝ともに、能力水準が高くなると選択率が低くなるという結果が得られた。また、いずれの錯乱枝も、能力低群—中群間でより選択率の差が大きくなっていったことから、能力水準の低い受検者と中程度の受検者を識別する錯乱枝であったことが示唆される。2種類の錯乱枝の選択率は類似した推移を示していたことから、両者は似た性質のものであった可能性も考えられる。

研究 6 では、一面的な記述・筆者の意図とのズレとともに、正答よりも出現頻度の高いカテゴリであった。多枝選択式項目を用いた本研究では、これらの特徴をもった要約文を錯乱枝として設定した場合に、能力群間の選択率の差が大きくなり、特に能力低群では能力中群よりも選択率が高いことが示された。いずれの錯乱枝も段落中の筆者の意図が適切に反映・表現されていない要約文であったと考えられるため、要点把握能力の高い受検者は、筆者の意図を適切に理解した上でこれらの錯乱枝を選択しなかったのに対し、要点把握能力の低い受検者は、錯乱枝として提示された要約文に筆者の意図が十分反映されているかどうかに関する判断を誤った結果、多くの割合でこれらの錯乱枝を選択していたものと考えられる。

なお、研究 6 の考察では、統合問題で取り上げたこれらの誤答の特徴が削除過程と密接に関連している可能性について論じた。例えば、研究 6 では一面的な記述の特徴が必要な要素の不足の一部となっている可能性を指摘し、錯乱枝作成の際に注意を必要とすること

を示唆した。本研究のデータを用いて、この点を補足的に確認してみると、2種類の錯乱枝の選択率の推移は類似したものではなく、必要な要素の不足錯乱枝を選んでいったことと、一面的な記述錯乱枝を選んでいったこととの間には相関がなかった( $r = -.027$ )。もし、一面的な記述錯乱枝の正誤判断の際に生じる認知過程が必要な要素の不足錯乱枝の認知過程と類似したものであれば、必要な要素の不足錯乱枝を選ぶ者が一面的な記述錯乱枝を選ぶという結果が得られるはずであるが、本研究ではそのような結果は得られなかった。この結果から、本研究で作成した一面的な記述錯乱枝については、必要な要素の不足錯乱枝を選ぶ際の認知過程とは異なるものであったことが示唆される。段落中の情報の取捨選択と、筆者の意図に沿った要約文の選択は、一定程度独立した認知過程である可能性が考えられる。

#### 4.4 第4章のまとめ

第4章では、1) 英語文章を要約する際の解答パターンを把握すること、2) 1の知見に基づいて錯乱枝の作成を行った際の選択率について、能力群別に検討すること、の2点を研究課題とした。研究課題1を解決するために行った研究6では、文章の要点を把握する中で生じる削除・一般化・統合の過程に特有の誤答の特徴が見い出された。また、研究課題2を解決するために行った研究7では、これらの誤答の特徴を反映した錯乱枝が受検者の能力差の識別に有効であることを示すことができた。以下では、本章における2つの研究課題に即して、研究6・研究7から得られた知見を整理する。

第1の研究課題「英語文章を要約する際の解答パターンの把握」に関して、研究6では以下の3点が明らかにされた。1つ目に、重要でない情報や冗長な情報を削除する過程では、重要でない情報を含む、必要な情報が不足しているなどの誤答の特徴が顕著に見られたことに加え、過度の削除や的外れ、要点把握以前の読解の誤りといった特徴も少数ながら確認された。2つ目に、具体的な表現を上位概念の語に置き換える過程では、文章中の具体的な表現を抜き書きしている、適切でない一般化を行っているなどの誤答の特徴が示された。3つ目に、段落中の情報を統合する過程では、段落中の記述の一部に重きを置いている、筆者の意図から逸脱した記述を含んでいるなどの特徴が見られた。

第2の研究課題「英語文章を要約する際の要約パターンに関する知見に基づき作成された錯乱枝の選択率に関する能力群別の検討」に関しては、いずれの錯乱枝についても能力群間で選択率の差が検出され、能力低群の選択率が中群に比べて高く、能力中群の選択率が高群に比べて高いという結果が得られた。削除問題における必要な要素の不足錯乱枝や、統合問題における筆者の意図とのズレ錯乱枝などに関しては、能力低群と中群の間で選択率の差が大きくなっていった一方で、一般化問題における具体的な記述の置き換え不足錯乱枝・不適切な一般化錯乱枝に関しては、能力低群と中群の間で選択率の差が小さく、

能力中群と高群の間で選択率の差が大きいという結果が得られた。

本章の2つの研究課題から得られた知見を通じて、設問の設定や選択枝の作成の工夫により、多枝選択式項目を用いて英語文章の要点把握能力を測定することが可能であることが示唆された。一般に、要点把握能力と言ったとき、記述式項目を用いて測定することが連想されやすいところである。ただし、記述式項目を用いて英語文章の要点把握能力を測定した場合には、採点基準の作成・修正や採点の一致度など、測定論的に克服すべき課題が多いため、受検者から記述式の解答データを収集した後のコストが極めて大きい。本章から得られた知見は、受検者の典型的な誤答の特徴を反映させた選択枝の作成により、多枝選択式項目を用いたとしても英語文章の要点把握能力を十分測定できる可能性を示唆するものである。一般に、多枝選択式項目は記述式項目に比べて信頼性が高く、採点が容易である反面、項目作成が難しいという課題を抱えている。本章が提供した知見は、錯乱枝のより効率的な作成を支援し、多枝選択式項目の欠点を補うことで、記述式項目よりもコストパフォーマンスのよい測定を可能にしようのものであると言える。

本章において受検者の典型的な誤答を捉えるために用いた方法はシンプルなものであり、記述式項目を用いて受検者の誤答の特徴のカテゴリを検討するという方法であった。この方法は、本論文がターゲットとする大規模テストだけでなく、教育現場で行われているテストにも適用可能な方法である。なお、受検者の誤答の特徴を詳細に捉えたい場合には、受検者に対する解答後のインタビューや発話思考法などを行うことが提案されており (Haladyna & Rodriguez, 2013)、テストの目的に応じて受検者の誤答の捉え方を工夫する必要が生じることも考えられる。



## 第5章

# 総括的討論

第5章では、5.1節において本論文で明らかにした研究知見を整理し、5.2節において本論文が提供した知見の意義と限界について述べる。次に、5.3節では本研究の今後の展望について述べ、5.4節で本論文を締めくくる。

### 5.1 本論文の知見のまとめ

本論文における7つの研究は、誤答の特徴を捉えるための研究、キーセンテンスの性質を確認するための研究、誤答の特徴を反映させた錯乱枝の選択率について検討するための研究の大きく3つに分類できる。このうち、誤答の特徴を反映させた錯乱枝の選択率について検討するための研究が本論文の中核であり、前者2種類の研究は、この研究を行うにあたっての情報収集・整理の役割を担っている。

本節では、誤答の特徴を反映させた錯乱枝の選択率について検討するための研究（研究2, 4, 5, 7）から得られた知見を整理する。研究2では、研究1において明らかにされた大学入学試験における英語文章読解問題の誤答の特徴を錯乱枝に反映させ、選択率に関する能力群別の検討を行った。研究4では、研究3において文章中でのキーセンテンスの位置関係に関する整理を行った結果をもとに、キーセンテンスの位置関係が否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の選択率に及ぼす影響について、能力群別の検討を行った。研究5では、キーセンテンスの特定を明示的に求める設問（下位レベル設問）と、キーセンテンスを特定した上で段落や文章全体の主旨について問う設問（上位レベル設問）を設定し、キーセンテンス中の語を錯乱枝中でも使用するかどうかを操作して、否定語錯乱枝・対義語錯乱枝・因果関係の取り違い錯乱枝の選択率に関する検討を能力群別に行った。研究7では、研究6において抽出された英語文章要約問題における誤答の特徴を錯乱枝に反映させ、選択率に関する能力群別の検討を行った。

本論文では、中核をなす4つの研究から、大きく以下の4点を明らかにできたものと考えられる。具体的には、1) 能力低群で初歩的な正誤判断の誤りを含んだ錯乱枝の選択率が

高かったこと、2) 能力低群で局所的な情報処理の誤りを含んだ錯乱枝の選択率が高かったこと、3) 能力中群で因果関係を取り違えた錯乱枝の有効性が示されたこと、4) 記述式問題から得られた誤答の特徴を錯乱枝に反映させるという手法が、英語文章の要点把握能力を測定する多枝選択式項目の作成に資する可能性があること、の4点であった。

第1に、能力低群では初歩的な正誤判断の誤りを含んだ錯乱枝の選択率が高いということである。例えば、文章中に記述されていないことを述べた錯乱枝や、文章中の記述に否定語・対義語などを付した錯乱枝（研究2）、キーセンテンスの特定を明示的に求める設問（下位レベル設問）でキーセンテンス中の語をそのまま使用した（重複条件）錯乱枝（研究5）、英語文章の要点把握を求める問題での必要な要素を欠いた要約文を提示した錯乱枝（研究7）などがこれに該当するものと考えられる。

上に挙げたそれぞれの錯乱枝に反映させた誤りはいずれも、認知過程上の初歩的なエラーとしてとらえることが可能である。例えば、文章中に記述されていないことを述べた錯乱枝は、キーセンテンスを特定するという処理自体をスキップしたときに生じる誤答である。また、キーセンテンス中の語をそのまま利用した上で、否定語・対義語・因果関係の取り違えなどの誤りを含んだ錯乱枝は、キーセンテンスと選択枝を見比べるという認知処理の不十分さから、キーセンテンス中の語を含んでいるかどうかで選択枝の正誤を判断しようとする認知過程から生じる誤答である。さらに、必要な要素を欠いた要約文を錯乱枝とした場合には、段落中の個々の記述内容に対する重要度の判断が不十分であるために生じる誤答である。これらの錯乱枝は、キーセンテンスを特定する、キーセンテンスと選択枝を見比べる、段落中の記述内容の重要度を判断するなど、英語文章読解・要点把握における最初のプロセスであると考えられる。能力低群において、上記のような誤答を含む錯乱枝の選択率が高かったことは、本論文における能力低群の受検者が、読解におけるこれらの初歩的なプロセスを十分に経ていなかった可能性が示唆される。本論文は、受検者の認知過程を想定した錯乱枝について検討を行うことにより、能力低群の受検者に見られる初歩的な誤答を具体的な形で明らかにし、それを反映させた錯乱枝が能力低群の受検者にとって魅力的であることを示すことに成功したのと考えられる。

第2に、能力低群では局所的な情報処理の段階の誤りを含んだ錯乱枝の選択率が高いということである。研究2・4・5を通じて、否定語錯乱枝や対義語錯乱枝など、キーセンテンス中の一部の記述を反転させた錯乱枝に関して、能力群間での選択率の差が顕著に見られた。特に、キーセンテンスが隣接する2~3文に集約されている状況（研究4）や、キーセンテンスの特定が明示的に求められている状況（研究5）では、能力低群と中群との間で選択率の差が大きくなっていた。これらの知見は、能力低群の受検者が英語文章中に含まれる文単位の理解でつまづいている可能性を示唆するものであり、問題作成者に多くのヒントを提供しうるものである。研究1で示した通り、わが国の大学入試として出題されている英語文章読解テストでは、文章中に記述されていないことがらを含んだ錯乱枝が多

く作成され、対義語など局所的な情報処理に焦点を当てた錯乱枝は相対的に少なかった。この種の錯乱枝の数が少なかったのは、英語文章中の文の局所的な情報処理の誤りに焦点を当てた錯乱枝の有効性が認識されていなかったためかもしれない。本論文の知見は、有効性が確かめられていなかった錯乱枝の有効性を実証的に示した点で、項目作成に対する示唆を提供したものと考えられる。

第3に、能力中群では因果関係を取り違えた錯乱枝が有用であるということである。代表的な結果として、一内容に集約されたキーセンテンスを踏まえる設問での因果関係の取り違い錯乱枝(研究4)が挙げられる。隣接する2~3文の内容に関する因果関係は、順接・逆接や原因・結果、時間的な前後関係など、局所的な因果関係ということになる。このような錯乱枝は、King et al. (2004)やLin et al. (2010)においてLevel 2(文章中に記述されている内容どうしの関係を取り違えた錯乱枝)に該当するものと考えられる。本論文の知見はこれらの先行研究と矛盾するものではなく、受検者集団の中で中程度の能力水準にある者の多くが選択していた。この知見は、錯乱枝を作成する際の知見として、実用性の高いものであると考えられる。

第4に、主に第4章の研究6・7から得られた知見であるが、記述式問題から得られた受検者の典型的な誤答を錯乱枝に反映させる手法が、より高次の読解能力の適切な測定に資する可能性があるということである。第1章でも述べた通り、英語文章の要点把握の際に用いられる削除・一般化・統合という認知過程は、文章中の個々の命題を適切に表象できていることを前提としたものであるため、個々の命題の理解に焦点を当てた場合よりも高次の読解能力を要するものと考えられる。このような高次の読解能力の測定にあたって、わが国では記述式問題を用いることが望ましいと考えられがちであるが、受検者の典型的な誤答の特徴を考慮して錯乱枝に工夫を施すことで、多枝選択式問題を用いることも有用であることが示された。

また、本論文の研究6・7で焦点を当てた英語文章の要点把握能力そのものに対しても、有用な示唆を提供したものと考えられる。本論文では、要点把握の側面に関連が深いと思われる文章要約モデル(e.g., Kintsch & van Dijk, 1978; Brown & Day, 1983)に立脚して検討を行った。削除過程では重要でない情報の混入や必要な要素の不足という特徴、一般化過程では具体的な記述の置き換え不足や不適切な一般化という特徴、統合過程では一面的な記述や筆者の意図とのズレという特徴が顕著に見られた。このような特徴は、受検者の認知過程に基づいた項目作成を行う際、測定対象とする能力を定義するにあたっての参考資料となりうるものである。

## 5.2 本論文における研究の意義と今後の課題

### 5.2.1 本論文における研究の意義

本論文では、多岐にわたる英語文章読解能力を題材として、受検者の典型的な誤答を多枝選択式項目の誤答選択枝に反映させた錯乱枝を作成し、その選択率を能力群別に検討する研究、およびその関連研究を7つ行った。本論文の7つの研究は、以下の3点において意義を有するものと考えられる。具体的には、1) 我が国の大規模かつハイステークスなテストにおける項目作成時の参考資料として機能する点、2) 認知診断を目的とするテストでの項目作成にも活用できる点、3) 受検者に対する実際の解答から誤答の特徴を見出し、錯乱枝に関する実証的知見を提供した点の3点である。

第1に、わが国の大規模かつハイステークスなテストにおいて、本論文の知見は項目作成者にとっての参考資料として機能する点である。第1章でも述べた通り、現在におけるわが国のテスト文化の下では、作成された項目の統計的性質を確認する目的の予備テストを行うことが難しいため、海外のテストのように、項目分析の結果に基づいて項目の修正を行うことができない場面が多い。わが国におけるこのような状況の下では、予備テストの分析結果に代わって、項目を作成する際の参考資料が必要となる。本論文では、受検者の典型的な誤答を反映させた錯乱枝の選択率を能力群別に検討し、それぞれの結果に対して認知的心理学的な解釈を試みている。本論文の知見は、予備テストを行った場合の項目分析結果に代わる役割を担い、どのような錯乱枝を作成すればよいかという問いに対するヒントを提供する機能をもっている。この点において、本論文には項目作成の現場をサポートするという実践的意義があると言える。

なお、大規模かつハイステークスなテストに限定したのは、本論文が全体にわたって認知的心理学の研究知見を基にしているためである。認知的心理学の知見には、人間の認知的な処理を一般的なモデルで表現しようとするものも多いため、認知的心理学の知見に基づいて項目作成を行う場合には、受検者個別の認知過程の優先度を下げ、多くの受検者に共通する認知処理の過程を項目に反映させていることになる。一方で、受検者個人の処遇をそれほど左右しないテスト(e.g., クラスルームテストなど)では、その受検者集団が特有にもつ特徴や置かれている状況などを加味した上で、項目作成を行う必要がある。このようなテストでの項目を作成する場合、本論文の知見が有効に機能する場面も存在すると考えられる一方で、受検者の特徴や状況によっては本論文の知見通りでない場面もありうる。したがって、ある程度限定された受検者集団に対するテストの項目作成を行う場合には、本論文の知見が適用できるかどうか、十分吟味する必要があるものと考えられる。

第2に、本論文の知見は選抜のためのテストだけでなく、認知診断を目的とするテストでも適用できる点である。認知診断とは、測定したい能力を構成する下位スキル(i.e., 認知



的アトリビュート)に焦点を当て、下位スキルごとに受検者の得意・苦手を把握することを指す(Leighton & Gierl, 2007). 本論文から得られた知見を適用すれば、複数の下位スキルが反映された項目を作成することも可能である.

第2章では、文章中の記述の有無・対概念の使用・言及対象の取り違えの3要因から錯乱枝の特徴とその選択率を明らかにした. これらの要因を、英語文章読解能力のうち比較的微細な箇所の読解を規定する下位スキルとして捉えた場合、これらを複数組み合わせるような項目を作成し、それに対応する錯乱枝として文章中の記述なし錯乱枝(文章中の記述の有無要因)・対義語錯乱枝(対概念の使用要因)などを配置すれば、このような下位スキルの習得状況を診断することが可能になるかもしれない.

また、第4章では、削除・一般化・統合の3側面から要点を把握する能力を定義し、受検者の誤答の特徴を錯乱枝の作成に活かして、選択率の様子を明らかにした. 例えば、削除と統合の両方を組み合わせることが必要な段落の要点の理解を求める問いを立てたい場合には、重要でない情報の混入錯乱枝(削除)や筆者の意図とのズレ錯乱枝(統合)を提示することなどが考えられる.

上記のように、本論文の知見を活用してテスト項目を作成すれば、認知診断モデル(cognitive diagnostic model)におけるQ行列(Q-matrix)のより適切な設定につながる可能性がある. 認知診断モデルのうちMC-DINAモデル(multiple-choice deterministic input noisy-and-gate model)では、各項目におけるそれぞれの選択枝と下位スキルを対応づけたQ行列を用いて下位スキルの診断を行う(de la Torre, 2009). 本論文から得られた錯乱枝の機能に関する知見は、認知過程と受検者の選択枝への応答との対応関係を示唆するものであることから、下位スキルと選択枝を対応づけたQ行列をより正しく設定することに貢献するものである. 具体的には、本論文で検討を行ってきたさまざまな誤答について、「文章中の記述に基づいて正誤判断するスキル」「否定語を含む記述を文章中のキーセンテンスと照らし合わせながら正誤判断するスキル」のような下位スキルを立て、そのような錯乱枝を含んでいけば1、そうでなければ0とコーディングしたQ行列を設定することが想定される. 本論文はテスト項目の妥当性を高めることを主眼に置いてきたが、受検者の得意・苦手の診断に活用可能な示唆を得た点でも、学術的意義・実践的意義ともに高いと考えられる.

第3に、項目作成の研究の中で、受検者の典型的な誤答を調査した結果に基づいて、錯乱枝に関する実験的な検討を行った点である. 先行研究の多くは、すでに実施されたテスト項目を用いて、錯乱枝の特徴を事後的に分類し検討を行ったものであった(e.g., Freedle & Kostin, 1991, 1993; Ushiro et al., 2007). この方法の下では、すでに出题されたテスト項目における錯乱枝の特徴しか検討対象にはならず、他の典型的な誤答が検討されないという問題点がある. 本論文では、受検者の解答データから典型的な誤答を抽出し、実験的に錯乱枝に反映させるという手続きをとった. この方法により、先行研究では取り上げられな

かった誤答も取り上げることが可能となり、より多くの錯乱枝に関する知見を提供することができた。本論文は、先行研究とは異なった視点で錯乱枝に関する知見を提供している点において、学術的意義があるものと考えられる。

また、本論文では一貫して、項目作成の現場ですぐに活用できる、即効性の高い知見の提供を目指し、受検者の認知処理上の誤りが明確になるように典型的な誤答を抽出した。このことが可能となったのは、受検者の解答データから誤答の特徴を抽出するという方法をとったためである。本論文が採用した方法は、錯乱枝を作成する際に必要な「受検者の具体的な誤り」の抽出を可能にした点において、実践的意義もあわせもつものと考えられる。

### 5.2.2 今後の課題

上に記した意義もある一方で、本論文には課題も複数存在する。本節では、以下の2つの観点から、本論文全体を通しての課題を整理する。第1に、さまざまな側面から見て知見の一般性に限界がある点、第2に、テストの具体的な目的や得点の利用方法などを考慮に入れた提案までには到達していない点が挙げられる。

第1の課題である知見の一般性の限界に関して、まず受検者集団が限定されていた点が課題である。本論文では、受検者集団の能力水準がある程度多様になるように配慮しているものの、センター試験などの国内の大規模試験における受検者集団と比べると、偏りがあるものと考えられる。本論文で得られた知見は、そのままであってもわが国の大規模試験における項目作成を改善するものと考えられるが、実際に大規模試験で活用するにあたっては、大規模試験の受検者集団を代表するような受検者の解答データを収集し、本論文の知見が一貫したものであることを示すことが必要である。

また、本論文の知見は予備テストを実施できる状況でも活用されうる一方で、受検者の能力水準とテスト項目の難易度の適合に関して、本論文の状況が多くの予備テストの状況とは異なっている点についての留意も必要である。本論文において、受検者に見られる誤答の特徴を適切に抽出できたり、能力の識別に寄与する錯乱枝を明らかにできたのは、受検者の能力水準に合った英語文章読解項目を用いたためであると考えられる。実際、本論文の多くの研究において、正答率が極端な値をとる項目は見られなかった。それに対して、多くの予備テストでは受検者の能力水準に合うテスト項目を用意できない状況も想定される。どのような誤答が魅力的に見えるかということは受検者集団に応じて異なることもあるため、場合によっては本論文の知見が活用できない可能性もある。本論文の知見は、受検者の誤答の特徴について見当がつかない状況において錯乱枝作成の指針としての機能を果たす可能性があるが、本論文の知見に基づいて作成したテスト項目と受検者の能力水準の不適合が明らかになった場合には、それぞれのテストがターゲットとする受検者

集団に対し、誤答の特徴を把握するなどの調査が改めて必要になるものと考えられる。

加えて、受検者数や項目数が本論文の知見の一般性を限定している可能性も考えられる。本論文では、最大でも460名程度の大学生を受検者としたが、10万名～100万単位位の受検者を抱える大規模試験に比べた場合、受検者数の少なさから生じる知見の不安定さには課題があるものと思われる。本論文中の研究では、解答時間や認知負荷に関する受検者への配慮、英語文章読解テストそのものがもつ性質などから、1種類のテスト冊子に収録した項目数に限界があり、多くの項目について検討することが困難であった。多くのテスト項目について検討することを可能にするためには、本論文のように心理学研究への参加という形式を採るのではなく、テスト実施機関におけるモニター調査や試行調査など、解答時間が長くなることや多少の認知負荷の高さに同意が得られる方法を用いることにより、多種多様な英語文章・テスト項目を検討対象とし、知見の一般性を高めることなども考えられる。

さらに、本論文では日本人を中心とする受検者に解答を求めているため、国際的に行われる英語文章読解テストでも適用可能な知見であるかどうかについては、追加の検討を必要とする。第二言語習得の領域では、母語の読解スキルが第二言語の読解にも影響を及ぼすことが知られている(e.g., Yamashita, 2002)ことから、選択枝の正誤判断に対しても母語が一定程度影響するものと考えられる。したがって、本論文の知見が複数の国と地域で行われる英語文章読解テストの項目作成を支援するものとなるためには、受検者の母語に応じて選択枝への選択状況が異なるかどうかを国際調査などにより確認し、本論文の知見を修正・再提供する必要があると考えられる。

第2の課題として、テストの目的や結果の利用方法を考慮に入れた提案までには至らなかったことが挙げられる。本論文では、受検者の能力群と錯乱枝の選択率との関連を示し、当該の錯乱枝が有効となるであろう状況や錯乱枝の配置の仕方をいくつか提案した。ただし、どのような錯乱枝を配置するかということは、テストの目的や結果の利用方法、受検者集団の特徴など、個別のテストの事情に応じて異なる。本論文では、主にハイステークスな大規模テストの妥当性を高め、テスト得点の解釈と利用(Kane, 2013)を想定して知見を提示したが、テスト得点に対する解釈の明確さと利用の適切さを目指すのであれば、それぞれのテストが抱える個別具体的な事情も勘案する必要がある。Embretson (2001)が提唱する認知デザインシステムにおいても、項目作成の最初の段階で、内容的妥当性の観点からみた測定のゴール(e.g., どの能力について、どのような情報を得たいか)と、基準関連妥当性の観点からみた測定のゴール(e.g., 測定する能力はどのような場面で発揮される能力か)についての検討が必要であるとされている。本論文では、それぞれのテストの具体的な事情を差し置いた一般的な知見を提供しており、それ自体でも有用な示唆を与えうるものであるが、テスト作成の現場に活用する場合には、本論文の知見の中から適切なものを取捨選択したり、追加の検討を行ったりする必要があると考えられる。

## 5.3 今後の展望

### 5.3.1 認知診断を目的としたテストにおける項目作成

本論文の知見は今後、認知診断 (cognitive diagnosis) に活用できる可能性が考えられる。認知診断を目的とするテストは、さまざまな点において選抜や資格認定のためのテストとは異なるものであるとされる (Leighton & Gierl, 2007)。なかでも、選抜や資格認定のテストと認知診断テストとの大きな違いは、結果の利用方法と能力の捉え方にあると言える。

選抜や資格認定を目的とするテストでは、合計得点を用いた能力値の指標 (e.g., 合計得点そのもの、潜在特性値  $\theta$  など) に基づいて、受検者の処遇に関する意思決定が行われる。このような状況下では、できるだけ誤差が含まれない得点を得ることに重きが置かれる。対して、認知診断を目的とするテストでは、測定対象の能力を構成する下位スキル別に受検者の得意・苦手を把握し、受検者自身の今後の学習改善や教師の授業改善に活用することが意図されている。

能力の次元数に関して、選抜や資格認定のためのテストでは、基準点の設定を容易にするため、合計得点の背後に一次元性を仮定することが望ましいとされる。それに対し、認知診断テストでは能力の多次元性が仮定され、複数の次元からなる下位スキル (認知的アトリビュート) が関連しあって受検者の応答が得られているという立場に立っている。

このように、選抜や資格認定のためのテストと認知診断のためのテストは、さまざまな点で違いが見られる。これらの違いを踏まえると、選抜に用いられるテスト項目と認知診断に用いられるテスト項目を作り分けることが必要になると考えられる。実際、認知診断のために用いられるテスト項目を他の目的で作成されたテスト項目で代替することは難しいとの指摘がある (Leighton & Gierl, 2007)。認知診断を行う上では、テスト項目が受検者の学習や教師の授業の改善に資する情報を引き出せるものである必要がある。

認知診断をより効果的に行うことのできるテスト項目を作成するためには、テスト開発全体を方向づける枠組みが必要となる。テスト開発の枠組みとして 2000 年代から注目を浴びてきたものに、証拠中心デザイン (evidence-centered design; ECD) がある。証拠中心デザインでは、論証モデルを背景とした Kane (2013) の妥当性理論に立脚するテスト開発の枠組みである (Mislevy et al., 2003; Almond et al., 2015)。測定対象とする下位スキルどうしの関係を生徒モデル (student model) として表現し、それぞれの下位スキルの習得状況を表した記述文 (claims) を作成した上で、記述文を論証するためにどのような証拠を受検者から収集するかを証拠モデル (evidence model) で明確にし、課題モデル (task model) に基づいて必要な証拠を収集できるテスト項目を作成することとなる。証拠中心デザインにおいては、テスト項目に下位スキルの習得状況を表した記述文を裏づける証拠としての役割が求められていると考えられる。

以上の整理を踏まえると、認知診断テストにおける項目を作成する際には、上に挙げた点についての詳細な検討が必要になると言えるだろう。このような課題が存在する中で、受検者の誤答を反映させた選択枝の選択率について検討を行った本研究の知見は、受検者の認知的特徴を把握することに寄与したものと考えられる。

### 5.3.2 自動項目生成の基礎的知見としての利用

本研究の知見は、自動項目生成 (automatic item generation) に向けた基礎的知見として活用することも可能である。自動項目生成とは、設問・正答選択枝・誤答選択枝のリストを整理した項目テンプレートに基づき、選択枝の組み合わせなどに関する制約の下でアルゴリズムを適用し、同一の仕様の項目を大量に自動生成するしくみのことである (Gierl & Haladyna, 2013; Gierl et al., 2012)。実際、医学の知識やスキルなどについて問う多枝選択式項目の自動生成を行った Gierl et al. (2012) では、1つの項目テンプレートから 1248 項目を自動生成できたことが報告されている。自動項目生成の技術を適用すれば、困難度や識別力が比較的等質な項目を大量に生成することが可能となるとされており (Gierl & Haladyna, 2013)、項目応答理論を適用したテストやコンピュータ適応型テスト (Computer Adaptive Testing, CAT) など、膨大な数の項目をもつ項目バンクの構築を支援する技術として期待されている。

自動項目生成において、アルゴリズムそのものはもちろん重要である一方で、測定対象の能力の定義から始まる項目テンプレートの作成手法も極めて重要である (Gierl & Haladyna, 2013; Gierl et al., 2012)。項目テンプレートの作成にあたっては、測定したい能力に関するレビューを行い、その結果に基づいて受検者の能力の様相を表現する記述文 (能力記述文) を生成すること、能力記述文に反映された能力に関する情報を効果的に得られるようなテスト項目形式を検討すること、テスト項目の特徴が受検者の解答に及ぼす影響を確認することなどの作成過程が必要である (Gierl & Haladyna, 2013)。自動項目生成を適用して困難度・識別力が比較的等質なテスト項目を大量に生成する技術の背後には、項目作成者がテスト項目を作成する場合と同様の、上に示したような検討過程が必要となる。

本論文が提供した一連の研究知見は、項目テンプレートの作成段階における参考資料として、有効に機能する可能性がある。英語文章を読み、多枝選択式項目に解答する際の受検者の認知過程を錯乱枝に反映させ、能力群別に選択率に関する研究知見を提供した本論文は、受検者の解答傾向に関する参考資料となるなど、項目テンプレートの効果的な作成に寄与するものであると考えられる。

本論文のように、測定対象の能力についてのレビューを行った上で、受検者の解答について検討した項目作成の研究はほとんど見られない。今後の項目作成研究では、自動項目生成を念頭に入れ、さまざまな観点から項目テンプレートの作成に寄与する知見を蓄積し

ていくことが望まれる。

## 5.4 結語

本章で紹介したように、近年では自動項目生成などの技術が開発され、項目作成にかかる時間と労力を軽減できる可能性が示されている。一方で、測定したい能力の定義やテスト形式の選択、設問や選択枝のテンプレートの作成にはどうしても人間が介在することになってしまう。Downing (2006a) や Gierl & Haladyna (2013), Haladyna & Rodriguez (2013) が指摘するように、項目作成にはアーティスティックな面があると言われ、テスト開発者や項目作成者の主観や信念が、テスト項目の統計的性質や信頼性、妥当性に大きな影響を及ぼしうる。教育測定の観点から見れば、このような状況は望ましくないと考えられるため、証拠中心デザインなどのテスト作成枠組みや、実証的研究に基づく項目作成ガイドライン等に沿った項目作成が提案されている。このような手法を用いた項目作成は、項目作成者個人に閉じた形ではなく、テスト開発に携わる者や先行研究の力を借りて行われる協働作業として位置づけられることになる。

本論文では通底して、受検者の誤答をどのように項目作成に活かすかという点に関する検討を行った。一見すると素朴な発想のように見えるかもしれないが、実際に項目を作成してみると、その素朴な発想を実現することがいかに難しいかがわかる。ただし、上に述べたように、項目作成が多くの関係者の協働作業であることを踏まえると、能力を測定される側の受検者本人にも項目作成に携わってもらうという考え方は極めて重要であると考えられる。本論文は、受検者の力を借りながら誤答選択枝を作成するという手法の有用性を示し、項目作成に有用となる実際的な知見を提供している点で、実際の項目作成を支援する大きな貢献をしたものと考えられる。

また、本論文は英語文章読解における受検者の誤答に着目することにより、英語文章読解能力の様相の一側面を浮き彫りにすることに寄与したものと考えられる。認知デザインシステムや証拠中心デザインの枠組みでは、認知モデルの構築・生徒モデルの構築を項目作成の前に行うことが一般的である。この工程では、測定対象の能力に関する詳細な検討が行われ、受検者にできること・できないことをレベル別に記述することが必要となる。本論文で明らかにされた英語文章読解能力の側面は、受検者のつまずきを記述するときの参考資料として有用に機能することが考えられ、テスト得点に対する適切な解釈基準の策定の礎を築くことに多大な貢献をしたものと考えられる。

## 引用文献

- Agresti, A. (2002). *Categorical data analysis*. 2nd ed. Hoboken, NJ: Wiley.
- 相澤 一美・石川 慎一郎・村田 年 (2005). 『大学英語教育学基本語リスト』に基づく JACET8000 英単語 桐原書店.
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. NY: Springer.
- Ascalon, M. E., Meyers, L. S., Davis, B. W., & Smits, N. (2007). Distractor similarity and item-stem structure: Effects on item difficulty. *Applied Measurement in Education*, 20, 153–170. doi:10.1080/08957340701301272
- Attali, Y., & Fraenkel, T. (2000). The point-biserial as a discrimination index for distractors in multiple-choice items: Deficiencies in usage and an alternative. *Journal of Educational Measurement*, 37, 77–86. doi:10.1111/j.1745-3984.2000.tb01077.x
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30, 441–465. doi:10.1177/0265532212473244
- Bock, D. R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51. doi:10.1007/BF02291411
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22, 1–14. doi:10.1016/S0022-5371(83)80002-4
- Brown, J., & Yamashita, S. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal*, 17, 7–30.
- Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, 47, 423–466. doi:10.1111/0023-8333.00016
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45–73. doi:10.1037/h0076248
- Carr, N. T. (2006). The factor structure of test task characteristics and examinee performance. *Language Testing*, 23, 269–289. doi:10.1191/0265532206lt328oa

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Cronbach, L. J. (1984). *Essentials of psychological testing*. 4th ed. New York, NY: Harper & Row.
- Cureton, E. E. (1957). The upper and lower twenty-seven per cent rule. *Psychometrika*, 22, 293–296. doi:10.1007/BF02289130
- 大学入試センター (2018). 平成 30 年度大学入試センター試験の志願者数 (確定) について【変更】 , [https://www.dnc.ac.jp/albums/abm.php?f=abm00011318.pdf&n=志願者確定\\_h30\\_0112.pdf](https://www.dnc.ac.jp/albums/abm.php?f=abm00011318.pdf&n=志願者確定_h30_0112.pdf). 2018 年 10 月 4 日閲覧
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33, 163–183. doi:10.1177/0146621608320523
- DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning*, 2, 1–23. {<http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4>}
- Downing, S. M. (2006a). Selected-response item formats in test development. In Downing, S. M., & Haladyna, T. M. (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M. (2006b). Twelve steps for effective test development. In Downing, S. M., & Haladyna, T. M. (Eds.), *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, 16, 486–514.
- Ebel, R., & Frisbie, D. (1991). *Essentials of educational measurement*. 5th ed. London: Prentice Hall.
- Educational Testing Service (2008). Validity evidence supporting the interpretation and use of TOEFL iBT scores. *TOEFL iBT Research Insight*, 4. [https://www.ets.org/s/toefl/pdf/toefl\\_ibt\\_insight\\_s1v4.pdf](https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf)
- Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, 11, 175–193.
- Embretson, J., Susan E Fand Gorin (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38, 343–368.
- Fagerland, M., Lydersen, S., & Laake, P. (2017). *Statistical analysis of contingency tables*. Boca Raton, FL: CRC Press.
- Freedle, R., & Kostin, I. (1991). The prediction of SAT reading comprehension item difficulty for expository prose passages. ETS Research Report RR-91-29.
- Freedle, R., & Kostin, I. (1993). The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main ideas, inferences and explicit state-



- ments. ETS Research Report RR-93-13.
- Gierl, M. J., & Haladyna, T. M. (2013). *Automatic item generation: Theory and practice*. New York, NY: Routledge.
- Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, *46*, 757–765. doi:10.1111/j.1365-2923.2012.04289.x
- Goodrich, H. C. (1977). Distractor efficiency in foreign language testing. *TESOL Quarterly*, *11*, 69–78. doi:10.2307/3585593
- Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, *42*, 351–373. doi:10.1111/j.1745-3984.2005.00020.x
- Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, *2*, 37–50.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Taylor & Francis.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*, 309–333. doi:10.1207/S15324818AME1503\_5
- 肥田野 直 (1972). 心理学研究法 7 テスト I 東京大学出版会.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*, 473–493.
- 一杉 武史 (2007). 私大英語の法則 ——中堅大学編—— アルク.
- 池田 央 (1973). 心理学研究法 8 テスト II 東京大学出版会.
- 池田 央 (2007). テストの科学 ——試験にかかわるすべての人に—— 株式会社教育測定研究所.
- 犬塚 美輪 (2006). 読解力の測定：プロセスとスキル 東京大学大学院教育学研究科 教育測定・カリキュラム開発（ベネッセコーポレーション）講座 2005年度研究報告書(2), 103–110.
- 犬塚 美輪・椿本 弥生 (2014). 論理的読み書きの理論と実践 ——知識基盤社会を生きる力の育成に向けて—— 北大路書房.
- 石井 秀宗 (2014). 人間科学のための統計分析 ——こころに関心があるすべての人のために—— 医歯薬出版.
- 石井 秀宗 (2018). 大学入試における共通テストの複数回実施は実現可能か ——日本のテスト文化やこれまで見送られてきた理由などからの検討—— 名古屋高等教育研究, *18*, 1–17.

- Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, *11*, 253–271. doi:10.1093/applin/11.3.253
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73. doi:10.1111/jedm.12000
- Karbalaei, A., & Rajyashree, K. S. (2010). The impact of summarization strategy training on university ESL learners' reading comprehension. *The International Journal of Language Society and Culture*, *30*, 41–53. 10.4304/jltr.3.1.135-140
- 加藤 健太郎 (2014). 教育系企業のアセスメント開発・運用実務における課題と展望 日本テスト学会第12回大会発表論文抄録集, 152–153.
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, *15*, 261–278. doi:10.1016/j.jslw.2006.09.006
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*, 17–24. doi:10.1037/h0057123
- Khaki, N. (2014). Improving reading comprehension in a foreign language: Strategic reader. *The Reading Matrix*, *14*, 186–200.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, *28*, 77–96.
- Kim, S. A. (2001). Characteristics of EFL readers' summary writing: A study with Korean university students. *Foreign Language Annals*, *34*, 569–580. doi:10.1111/j.1944-9720.2001.tb02104.x
- King, K. V., Gardner, D. A., Zucker, S., & Jorgensen, M. A. (2004). The distractor rationale taxonomy: Enhancing multiple-choice items in reading and mathematics. *Pearson Assessment Report*, July 2004, 1–14.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. doi:10.1037/0033-295X.85.5.363
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.
- Kuramoto, N., & Koizumi, R. (2016). Current issues in large-scale educational assessment in Japan: Focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy and Practice*, 1–19. doi:10.1080/0969594X.2016.1225667
- Leighton, J. P., & Gierl, M. J. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. New York, NY: Cambridge University Press.
- Lin, J., Chu, K. L., & Meng, Y. (2010). Distractor rationale taxonomy : Diagnostic assessment of reading with ordered multiple-choice items. *Paper presented at the annual meeting of American*

- Education Research Association*, 1–15.
- Lo, J. J., Yeh, S. W., & Sung, C. S. (2013). Learning paragraph structure with online annotations: An interactive approach to enhancing EFL reading comprehension. *System*, *41*, 413–427. doi: 10.1016/j.system.2013.03.003
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*, 741–749. doi:10.1037//0003-066X.50.9.741
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design., No. ETS Research Report RR-03-16.
- Morimoto, Y. (2008). Attractive distractors in sentential gap-filling tests: With a focus on proficiency levels. *Annual Review of English Language Education in Japan*, *19*, 211–220.
- 邑本 俊亮 (1992). 要約文章の多様性 教育心理学研究, *40*, 213–223.
- 野口 裕之・大隅 敦子 (2014). テスティングの基礎理論 研究社.
- Oded, B., & Walters, J. (2001). Deeper processing for better EFL reading comprehension. *System*, *29*, 357–370. doi:10.1016/S0346-251X(01)00023-9
- Ozaki, K. (2015). DINA models for multiple-choice items with few parameters: Considering incorrect answers. *Applied Psychological Measurement*, *39*, 431–447. doi:10.1177/0146621615574693
- Princeton University (2015). WordNet Search 3.1. <http://wordnetweb.princeton.edu/perl/webwn>. 2015年12月30日閲覧
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*, 3–13.
- Rodriguez, M. C., Kettler, R. J., & Elliott, S. N. (2014). Distractor functioning in modified items for test accessibility. *SAGE Open*, *4*. doi:10.1177/2158244014553586
- 清水 真紀 (2005). リーディングテストにおける質問タイプ——パラフレーズ・推論・テーマ質問と処理レベルの観点から—— *STEP Bulletin*, *17*, 48–62.
- Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*, *9*.
- 寺尾 尚大・石井秀宗 (印刷中). 英語文章要約パターンの教育測定学的検討 ——削除・一般化・統合のプロセスに着目して—— 日本テスト学会誌, *15*.
- 寺尾 尚大・安永 和央・石井 秀宗・野口 裕之 (2015). 能力別にみた錯乱枝の効果に関する検討——多枝選択式の英語文章読解テストを用いて—— 日本テスト学会誌, *11*, 1–20.
- 寺尾 尚大・石井 秀宗・野口 裕之 (2017). キーセンテンスと錯乱枝の語の重複・設問タイプが錯乱枝の選択率に及ぼす影響 ——英語文章読解テストを用いた実証的検討—— データ分析の理論と応用, *6*, 63–82.

- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10, 159–169. doi:10.1080/0950069880100204
- Ushiro, Y., Morimoto, Y., Hijikata, Y., Nakagawa, C., Watanabe, F., Kai, A., Shimizu, H., Koga, T., Ohno, M., Umehara, C., Hamano, Y., Narumi, T., Tao, N., Shimada, S., Kitayama, M., Suzuki, R., Kurisu, Y., & Gomi, N. (2007). What makes distractors plausible in multiple-choice reading tests? *The Japan Language Testing Association*, 10, 56–67.
- Whitely, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197. doi:10.1037/0033-2909.93.1.179
- Wilson, M. (2005). Construct modeling: The “four building blocks” approach. In Wilson, M. (Ed.), *Constructing measures: An item response modeling approach*. Taylor & Francis.
- 山口 一大 (2016). 認知診断モデルにおける Q 行列の誤設定が診断精度に与える影響 —— 認知の階層構造を考慮した場合の検討—— 日本テスト学会誌, 13, 17–32.
- Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, 25, 81–95. doi:10.1111/1467-9817.00160
- 安永 和央・石井 秀宗 (2012). テストにおける設問の問い方が回答傾向に及ぼす影響 —— 国語テストを用いた実証研究—— 教育心理学研究, 60, 296–309.
- 安永 和央・齋藤 信・坂本 雄士・石井 秀宗 (2013). 回答欄の字数制限が能力評価に及ぼす影響 —— 国語読解テストの出題傾向に着目して—— 日本テスト学会第 13 回大会発表論文抄録集, 130–131.
- Zhao, R., & Hirvela, A. (2015). Undergraduate ESL students' engagement in academic reading and writing in learning to write a synthesis paper. *Reading in a Foreign Language*, 27, 219–241.

## 謝辞

本論文の執筆にあたり、多くの方のご指導とご支援を賜りました。

まず、指導教員である名古屋大学大学院・石井秀宗先生には、学部生のときから現在に至るまで、厳しいながらも温かいご指導をいただきました。小生を教育測定学の道に導いてくださったのも石井先生であり、問題作成者に対する示唆を常に考える姿勢やデータ分析に対する姿勢など、教育測定学を専門とする研究者としてのあり方を教えていただきました。また、時に礼を失することのある未熟な小生に対して、他者を尊重することの重要性、礼節をわきまえることの大切さを教えていただきました。心より御礼申し上げます。名古屋大学名誉教授・野口裕之先生には、研究に対する具体的なご助言をいただいたことはもちろん、統計モデルの数理的理解の重要性、時間をかけてじっくり研究課題に向き合うことの大切さを教えていただきました。深く感謝いたします。名古屋大学大学院・光永悠彦先生には、本稿の極めて肝要な部分に対するご指導をいただきました。厚く御礼申し上げます。名古屋大学大学院・高井次郎先生には、研究遂行にあたってデータ収集の機会を豊富に提供してくださるなど、本稿の執筆をさまざまな観点からご支援いただきました。深く感謝いたします。

次に、本稿中の研究の実施に対してご支援いただきました先生方に感謝いたします。名古屋大学大学院・清河幸子先生には、主として SONA システムを用いた研究参加者募集の側面から、多大なご支援をいただきました。御礼申し上げます。また、九州大学・安永中央先生には、データ収集の機会を多く提供して下さり、小生の研究を温かくご支援いただきました。深く感謝いたします。植草学園大学・金子功一先生、愛知淑徳大学・平島太郎先生、名古屋大学大学院・玉井颯一研究員にも、データ収集の機会を提供していただきました。ここに御礼申し上げます。

さらに、本稿中の研究に対して多くのご助言をいただいた先生方に御礼申し上げます。名古屋大学名誉教授・村上隆先生には、大局的な観点から本稿中の研究へのご助言をいただきました。厚く御礼申し上げます。南山大学・解良優基先生、愛知学泉大学・坪田祐基先生には、それぞれの研究の位置づけや立場、社会に対するメッセージなどについて、多くのご助言をいただきました。深く感謝いたします。

現所属の大学入試センターでは、本稿の執筆と研究プロジェクトの遂行の二足の草鞋を履く小生に対し、多くの先生方の応援をいただきました。大学入試センター・大久保智哉先生には、本稿の執筆を最優先に取り組めるよう、常に温かく応援していただきました。大変感謝しております。大学入試センター・大塚雄作先生、大学入試センター・内田照久先生には、本稿の執筆状況を気にかけていただきました。深く感謝いたします。

最後に、学位論文の執筆を主として精神的な面から支えてくれた妻・香那子に感謝の意を表します。