

# テスト項目と英文読解ストラテジーの関係

## —正誤答時の視線データを基に—

愛知県／名古屋大学大学院在籍・日本学術振興会特別研究員 吉川 りさ

申請時：愛知県／名古屋大学大学院在籍

### 概要

本研究は、読解力テスト解答中における認知プロセスを明確にするため、日本語を母語とする英語学習者が読解テストに解答する際の眼球運動を計測し、(1)テスト項目の認知的妥当性の検証と、(2)解答者の内的要因と認知プロセスの関連を調べた。具体的に、大学(院)生が英検の読解問題を解答する際の眼球運動の計測と、アンケート・インタビュー調査を実施した。主な結果は以下のとおりである；(1)項目正答者は、誤答者に比べて、解答時のテキスト注視時間が短く注視回数が少ないことから、解答該当箇所をより迅速にかつ確実に認識している；(2)テスト項目正答に至るまでの認知プロセスには読み手のメタ認知ストラテジーが関与している；(3)解答時の認知プロセスを解明する上で、研究方法論問のトライアングレーションは有効に機能する。(1)を通して、英検問題項目への認知的妥当性が示された。これらの結果は、新たな視点からテスト評価と英語力評価の実現可能性を示唆している。

## 1

### はじめに

読解力テストに解答する際、テキスト情報にアクセスしてから解答を終えるまでの時間軸では、さまざまな認知プロセスが経られている。例えば、該当言語の文字ないし書記素単位での視覚情報の認識や音韻符号化（視覚的な言語情報から音韻表象を形成すること）、単語認知、複数語から成る意味のまとめ（チャンキング）の認識・理解、単文単位における統語処理、文章・文脈単位における談話理解や

推論などが挙げられる（Grabe, 2009）。また、テスト解答時には、テキストや選択肢に含まれるキーワードの同定や、解答該当箇所の意味・統語処理のような問題解答に特化した認知プロセスも同時にかかわっていると考えられる（Khalifa & Weir, 2009）。この認知プロセスは、テキストやテスト項目の種類・形式のようなテスト解答者にとっての外的要因と、読解に関するメタ認知ストラテジーや習熟度などの内的要因とが複雑に関与しており、個人間でさまざまに異なる。こういった認知プロセスの相違は、項目に対する難易度あるいは解答所要時間などの違いにつながり、テスト得点として表面化するが、得点が実際に意味するものを理解するためには、各項目への解答中に、解答者が何を考え、どのような解答法で答えを導き出したかという認知プロセスそのものを解明する必要がある。

そのため、従来の第二言語習得研究（ここでは、読解研究に焦点を当てる）では、Thinking aloud（タスク遂行時に思いついたことを口に出させる方法）や内省（タスク後に活動時の振る舞いを報告させる方法）などの言語報告法や質問紙調査法などを用いて、直接観察では見られない解答者の内面をとらえようとする試みがされてきた（例：Sasaki, 2000; Weir, Hawkey, Green, & Devi, 2009）。近年になると、測定技術が発展し、より忠実かつ客観的データを提供できる眼球運動計測も使用され始めてきている（Bax, 2013; Bax & Weir, 2012; Kruger & Steyn, 2013; Leung, Sugiura, Abe, & Yoshikawa, 2014）。読解研究における眼球運動計測では、読み手が、どの箇所（例：単語・段落）を、いつ、どの程度、どのように注視（視点が止まること）したかをミリ秒単位で

記録できるため、テキストを読み始めて解答を終えるまでのパフォーマンスを観察することが可能となる。これらの研究手法は、テスト解答時の認知的妥当性（テスト解答時に経るプロセスは適切かどうか）における有用性が示唆されてきたが（例：Bax, 2013; Bax & Weir, 2012; Weir, et al., 2009）、先行研究で残されている問題点は、テスト項目の種類だけに焦点を当て、個人の読解ストラテジーといった内的要因を考慮していない点である。この問題を解決するには、研究方法論問手法のトライアングレーション（ある1つの研究手法で収集されたデータを、異なる手法を用いて多角的にとらえるプロセス）を通して、多角的な角度からテスト解答時の認知プロセスを検討することが必要になる。

そこで本研究は上記の問題を解決するため、以下の2つの側面に焦点を当てて調査を行う。1つは、テスト項目の認知的妥当性の検証であり、もう1つは、解答者の内的要因と認知プロセスの関連である。本稿の構成としては、まずテスト解答時の認知的妥当性の概念を概観し、本研究の背景となるBax (2013) とその先行研究の知見を整理する。次に、これらの研究の限界と課題を指摘した上で、本研究が実施した2つの調査結果を報告する。1つ目は、英検の読解力テスト解答時の解答者の認知的妥当性を検証するため、テスト解答時の眼球運動データと各テスト項目の特性との関連を調べる。具体的には、項目ごとの正答者と誤答者の眼球運動データを比較し、正誤反応を弁別する要因を探る。2つ目は、項目に解答反応するまでに経る個人間の認知プロセスの違いと、その眼球運動との関連を調査するため、

個人間における読解に関するメタ認知ストラテジーとテスト項目解答時の眼球運動との関連を調査する。

## 2 先行研究

### 2.1 認知的妥当性

読解力テストに解答するには、各テスト項目が問うていることを理解し、それに対応した解答プロセスを経る必要がある。仮に、問われていることに準じた反応をしていなければ、誤答につながる可能性が高くなる。これを認知的妥当性という。読解力テストを構成する項目は、認知プロセスの種類によって表1のように階層化されると言われ（Khalifa & Weir, 2009）、学術的・専門的なリーディング力を測定するようなテストでは、さまざまな認知プロセスのレベルを問う項目が混在し、解答者のあらゆる能力をカバーしている必要がある（Bax & Weir, 2012）。Weir, et al. (2009) は、高水準の学術的な英語能力を測定する目的で開発され、国際的な規模で使用されているようなテストは、広範囲の認知プロセスを問う項目から構成されているという仮説に基づき、International English language testing system (IELTS) のアカデミック・モジュールのリーディングテストを実施し、解答時の読解プロセスに関するアンケート調査を実施した。そして、表1の1から5の認知プロセスを問う項目を分析対象として設定し、それらの項目に対する認知的妥当性を検証した。本研究の基となるBax (2013) は、このWeir, et al.

■表1：読解テスト解答時にかかわる認知処理のレベル (Bax, 2013, p.443)

| 活動レベル          | 読解力テスト解答時に用いられる典型的な認知プロセス                   | 対象単位    |
|----------------|---|---------|
| 1 語彙：単語同定      | 設問と文章の両方で使われている単語を特定する                      | 単語      |
| 2 語彙：同義語と品詞の同定 | 同義語や反意語、他の関連する語を特定するために、該当語の意味あるいは品詞情報を利用する | 単語      |
| 3 文法/統語        | 解答を特定したり判別したりするために、文法知識を使用する                | 節/文     |
| 4 命題的知識        | 語彙や文法の知識を使用して文の意味を正確にとらえる                   | 文       |
| 5 推論           | 字面の理解を超えて、テキストが含意する情報を的確に推論する               | 文/段落/文章 |
| 6 メンタルモデルの形成   | テキスト要素間の情報を関連づけて、より大きなテキスト表象を形成する           | 文章      |
| 7 テキストの機能構造の理解 | テキストジャンルを理解し、そのジャンル特有のテキスト構造の特性や、その目的を特定する  | 文章      |

(注) この表は、Khalifa and Weir (2009) の考えを基に Bax (2013) が作成した表 1 (ibid. p.443) を筆者が和訳したものである。

(2009)の研究デザインを参考にし、研究手法をアンケート調査でなく、眼球運動計測法に変えて、IELTS テスト項目に対する認知的妥当性を検証した。次節で Bax の調査内容を概観する。

## 2.2 Bax (2013) の調査内容

Bax (2013) の研究目的は、Weir et al. (2009) とは異なる研究手法——眼球運動計測——で、読解力テスト解答中の読み手の認知プロセスを観察し、各項目に対する認知的妥当性を検証することであった。そこで Bax は、マレーシアの大学に通う大学生71名（母語はさまざまに異なる）を対象に、IELTS アカデミック・モジュールのリーディングテストの空所補充問題（項目数5）と内容理解問題（項目数6）の各1題を読解力テストとして実施し、そのうちランダム抽出した38名のテスト解答中の眼球運動を計測した。この研究のもう1つの着眼点は、メタ認知ストラテジーであった。読み手のストラテジーの知識の有無と、目的に応じたストラテジーの使い分け（コントロール）は、優れた読み手（good readers）とそうでない読み手（poor readers）を弁別する1つの要因であり、認知負荷の高い読解力テストを第二言語で解答する場合、ストラテジーを駆使して認知資源を配分させた効率良い読みは、より良いパフォーマンスにつながると考えられている（Carrell, 1989）。そのため Bax は、眼球運動計測後の参加者にインタビュー調査も実施し、テスト解答中の内省報告をさせ、読み手のストラテジーの有無とその種類を検討した。

Bax の研究で調査対象となったテスト項目は、表1の1, 2, 3, 5に当たる認知プロセスであったが、それらの項目に対して正答者・誤答者との間で眼球運動計測データに有意な差が見られたのは、5項目（それらが問う認知プロセスは、表1の1, 2, 3に該当）であった。分析に使用した眼球運動計測値の指標は、テキスト総注視時間および総注視回数と、テキスト内の解答該当箇所への総注視時間および総注視回数であった（眼球運動の理論的枠組みは次節2.3で触れる）。そして各項目における正答者・誤答者の眼球運動データの比較とインタビュー結果に基づき、Bax は以下の結果を導いた；(1) 項目誤答者はテキスト内の解答該当箇所を効率よく特定することができず、文の前後を行き来した読みをする；(2) 項目正答者は、スキミング（テキストに素早く目

を通し、特定の情報を得る）をしているので、読解時間が短い；(3) 項目正答者は単語レベル（設問とテキスト内にある同義語やキーワードの同定）で高い能力を発揮する。また、(1)から(3)の結果から、眼球運動計測は、テスト項目に対する認知的妥当性を検証する上で有効な手法の1つであると言える述べた他、IELTS の読解力テストは、認知的妥当性が示された項目で構成されているという結論を出した。これまでの読解研究で、読解プロセスの解明を目的に眼球運動計測は使用されていたものの、テスト分野への応用がされてこなかった。その点を考慮すると、Bax の研究成果は、いくつかの問題点は残されているもの（2.4に詳述）、読解の認知プロセスに関心を持つ研究者や、教師、テスト開発者などに有益な示唆を与えている。

## 2.3 読解研究における眼球運動計測

本研究は主に眼球運動計測を用いて、読解力テストの解答時の認知プロセスを検証する。読解において視線がどのように動くかについては、認知・心理学の分野でこれまで多くの研究が行われてきており、主な基礎知見は以下のようにまとめられている。

読解において、文やテキストなどの刺激対象に視線が留まることは「注視」(fixation)、視線がある箇所から別の箇所に移動することは「サッカード」(saccade)と呼ばれる。Rayner (1998, 2009) によると、読解に際して、英語母語話者の1つの注視の平均的な長さ、すなわち、平均的な注視時間は約200ミリ秒から250ミリ秒程度である。サッカードの平均的な距離は、約7～9文字分である。サッカードの方向は、英語の読みにおいては主に左から右へととなっているが、約10～15%のサッカードは、右から左へ行われており、こうしたサッカードは逆行 (regression) と呼ばれる。また、読解中に必ずしもすべての単語には注視が置かれるとは限らず、約25～30%の単語は注視されずに読み飛ばされている。

こうした読解時の眼球運動は、Rayner (1998, 2009) で述べられているように、読む材料またはテキストに関する言語的特徴によって異なる。例えば簡単なテキストを読む場合は、難しいテキストを読む場合と比べて注視および逆行の回数が少なく、注視時間も短い。また、テキストの難易度のみならず、読み手の読解力や言語能力によっても読解時の眼球運動は異なる。

このように、読解時の眼球運動は読み手がテキストに対して払う注意と関連すると考えられている (Rayner, 1998)。近年、第二言語習得および言語教育の分野においては、眼球運動計測について興味を持つ研究者が増えてきている (Dussias, 2010; Frenck-Mestre, 2005; Roberts, 2012; Roberts & Siyanova-Chanturia, 2013)。しかしながら、前節で述べたように、こうした眼球運動計測を用いた第二言語および外国語の研究の中で、言語テストやテストニングの分野に焦点を当てた研究例は少数であるのが現状である。

## 2.4 先行研究の改善点およびそれらの解決案

筆者が知る限り、Bax and Weir (2012) と Bax (2013) がこれまでに、読解力テスト解答時の認知プロセスを検証した唯一の実証研究である。眼球運動計測をテストニングの分野に取り入れて、テスト開発やテスト評価の側面に有益な示唆をもたらした点は評価に値すると考えられるが、改善点がいくつか残されている。以下にそれらを列挙するとともに、本研究の解決案を述べていく。

### 2.4.1 実験デザインへの解決案

1つ目の改善点は、実験デザインである。具体的に言えば、1つのテキストに複数の項目が含まれているテストが問題として用いられている点である。実際のテスト実施場面に即した環境下で実験を行うことも1つの可能性として考えられるが、このようなデザイン下で実験を実施してしまうと、先行項目の解答経験が、後続の項目解答時の認知プロセスを抑制し、本来観察しようとする解答プロセスを的確にとらえられない可能性が十分考えられる。そこで解決案として本研究では、1つのテキストには1つのテスト項目のみを設けるデザインを採用し、各項目への直接的な反応を観察することにする。

### 2.4.2 テスト問題および項目数への解決案

2つ目の改善点は、テキストと項目の数である。Bax (2013) では、2つのテキストのみを実験材料として使用したが、これではトピック依存 (与えられたトピックに対する個人間の背景知識や興味の度合いの相違が解答に影響すること) の問題が考えられる。また、テストの信頼性の観点から、項目数11

はやや僅少であると思われる<sup>(注1)</sup>。トピックの影響を最大限排除するためには、さまざまなトピックを扱うテキストを複数用意する必要があり、そして、十分な信頼性を得るためには項目数をできる限り増やす必要がある。そこで本研究は、材料として用いるテキスト数を大幅に増やす策をとる。しかしながら、テキスト数の増加は、物理的に読むべき単語数の増加につながり、解答者に心身的な負担をかける恐れがある。その可能性を少しでも低減させるために、テスト選択時の条件として、テキストは100語前後で構成されるものに限定する。

### 2.4.3 質的調査への解決案

3つ目の改善点は、研究方法論間トライアングレーションである。Bax (2013) では、眼球運動計測では観察できない読み手の内面 (実際に解答時に感じたこと・気付いた点のような思考プロセス) を考慮し、眼球運動データを補完する手段としてインタビュー調査を用いた。眼球運動計測とインタビュー調査の2つの研究手法の統合は、互いのデータを補強し合う点で有効に機能すると考えられるが、問題点として残るのは、インタビューデータの質である。参加者によっては、インタビュー調査に慣れていない、恥ずかしい、答えたくない、などといった要因で、実験側が欲する情報を得られない場合がある。さらに、Bax は、テスト解答中の思考プロセスの内省のみに焦点を当てているが、テスト解答には、解答中の思考プロセスの他に、読解に対する習慣的なストラテジー (普段、読解をするときはどのような読みを行っているか) も関連していると考えられる。なぜなら、テスト実施時点の参加者の読解力は、これまでの読解・解答経験やその際に会得したストラテジーとその使用経験量などの要因が深く関連していると考えられるためである。そこで本研究では、眼球運動データを補完するために、インタビュー調査でテスト解答中の思考プロセスを検討すると同時に、アンケート調査で個人の持つ習慣的な読解ストラテジーを調査する。アンケート調査を用いることで、実験側が被験者間で同質の内容を、確実に、短時間で、間接的に (実験者と参加者が対話を通して直接接触することなく)、得られる長所がある。眼球運動計測とインタビュー調査、アンケート調査の3つの研究方法論的トライアングレーションを通して、テスト解答時に各解答者がテ

キストを、いつ・どこを・どれほど・なぜ、そのように読むかを把握することができ、読解時の認知プロセスを解明する上で有益な示唆をもたらすことが可能となる。

#### 2.4.4 本研究の調査内容

上記の先行研究の知見を踏まえて、本研究が設定するリサーチクエスションは以下のとおりである；

1. 眼球運動計測は読解テスト解答時の読み手の認知プロセスを明らかにするのか
2. 項目正答者と誤答者の間でどのように視線の動きが異なるのか
3. 読解に関するメタ認知ストラテジーの使用は読み手の眼球運動をコントロールするか

リサーチクエスション1と2に取り組むために、Bax (2013) に倣い、テスト解答時の解答者の眼球運動を計測し、項目誤答者と正答者の眼球運動計測データの比較を通して、項目に対する認知的妥当性を検証する。また、リサーチクエスション3に取り組むために、テキストを読み始めてから項目に解答するまでに経るメタ認知ストラテジーの使用とその種類をアンケート調査で検証し、その結果がテスト解答時の解答者の眼球運動とどのような関連があるのかを調査する。

## 3 方法

### 3.1 研究参与者

本研究の参与者は、名古屋大学の学部生40名および大学院生2名の計42名（男性13名、女性29名）であり、年齢幅は18~24歳（最頻値は20）であった。参与者は少なくとも6年間は日本での教育機関で英語を学習しており、大学（院）での専攻は、文学、教育学、工学、農学、医学、理学などさまざまであった。調査時の彼らの英語力（自己申請）は、TOEIC平均649点（SD=154.10）であった。すべての参与者は裸眼あるいは矯正により健全な視力を有していた。

### 3.2 マテリアル

#### 3.2.1 読解力テスト

読解時の眼球運動を計測するために、本研究では、

英検準2級と2級の読解セクションにおける長文読解問題32テキストを選定・使用した。テキストは、2011年度第3回から2014年度第1回の間で実施された計8回分の過去問題集から抽出された。準2級の問題からは第4問AとBの8テキスト、および第5問AとBの8テキストを、2級の問題からは第3問AとBの8テキスト、および第4問AからCの8テキストを選定した。なお、2.4.1および2.4.2節のBax (2013) の改善点で説明したとおり、問いへの反応（該当問題の解答時の眼球運動）を直接的に観察し、研究参与者間での比較を可能にするため、また、テキストのトピック依存の影響を排除し項目数を増やすため、テキストの使用箇所は第1段落のみとした。第1段落に対応するテスト項目は、各問題の間1であったため、それを使用し、本研究が使用するテスト問題はすべて、1テキスト（段落）1テスト項目で構成されるよう配慮した。

英検の長文読解問題の出題テキストには手紙形式や電子メール形式、説明文形式から成る内容理解問題と、説明文形式から成る空所補充問題があり、文体や問題形式の種類が多様であるが、このようなテスト形式の相違も個人の認知プロセスに影響すると考えられるため、特定の文体や問題形式に依存したテキスト選定は行わなかった。しかし、上記の影響をより鮮明に観察できるように、テキストの文数・単語数・テキスト数は級間でなるべく均等になるよう配慮を行った。本研究で用いた空所補充および内容理解問題のテキストの特徴（文数・単語数・テキスト数の平均値）は、表2および表3を参照されたい。

■ 表2：空所補充問題における使用テキストの特徴

| 級  | 形式  | 文数   | 単語数   | テキスト数 |
|----|-----|------|-------|-------|
| 準2 | 説明文 | 5.62 | 82.37 | 8     |
| 2  |     | 5.43 | 92.87 | 8     |

■ 表3：内容理解問題における使用テキストの特徴

| 級  | 形式     | 文数   | 単語数   | テキスト数 |
|----|--------|------|-------|-------|
| 準2 | 手紙・メール | 5.75 | 68.25 | 4     |
| 2  |        | 4.5  | 75.25 | 4     |
| 準2 | 説明文    | 4.75 | 66.25 | 4     |
| 2  |        | 4.75 | 87    | 4     |

### 3.2.2 読解に関するメタ認知ストラテジーに関するアンケート

読解力テスト解答時の眼球運動と読み手の習慣的な読解ストラテジーとの関係性を検証するため、そして2.4.3のBax (2013) の3つ目の改善点として挙げた参与者間のインタビューの質の違いを補完するため、30項目から成る Metacognitive Awareness of Reading Strategies Inventory (MARSИ: Mokhtari & Reichard, 2002) を実施した。MARSИは、英語で書かれた教材(例:教科書や参考書, 図書)を普段どのように読んでいるかを, Global Reading Strategies (「テキストを読むときにどう考えるか」というリーディングに対する意識に関する項目), Problem-Solving Strategies (「テキストの内容が難

しい」と感じるときにとる行動に関する項目), Support Reading Strategies (「メモを取る, 辞書を使う」のようなリーディングを促進させるためにとる行動に関する項目)の3つの観点から測定する質問紙調査である。回答法は, 「1 全くしない」から「5 ほとんどの場合(ほぼ必ず)する」までの5件法を用いて行われた。原文は英語であったため, まず筆者が日本語文へ翻訳し, その翻訳文を英語学専攻の大学院生2名に添削を依頼し, 日本語文への翻訳時に生じる誤翻訳の影響を最大限回避・低減した。さらにアンケート調査時においては, 原文と翻訳文の両方を提示し, 参与者が理解しやすい言語で回答をするよう指示した。各観点の質問項目とそれに対応した翻訳文は表4のとおりである。

■表4: MARSИ (Mokhtari & Reichard, 2002) の質問項目 (原文および筆者による翻訳文)

| 観点   | 原文  | 翻訳文                                     |
|------|---|---|
| *GRS | I have a purpose in mind when I read.   | 英文を読むときは, 何かしら目的を持っている。                 |
| *SRS | I take notes while reading to help me understand what I read.                     | 英文を読むとき, 内容が理解しやすくなるようにメモを取っている。        |
| *GRS | I think about what I know to help me understand what I read.                      | 読んでいる内容が理解しやすくなるように, すでに知っていることを考えている。  |
| *GRS | I preview the text to see what it is about before reading it.                     | 読む前に, 何が書かれているかあらかじめ英文に目を通す。            |
| *SRS | When text becomes difficult, I read aloud to help me understand what I read.      | 英文が難しい場合, 理解しやすくなるよう声に出して読む。            |
| *SRS | I summarize what I read to reflect on important information in the text.          | 英文に書かれている要点をまとめている。                     |
| *GRS | I think about whether the content of the text fits my reading purpose.            | 英文の内容が, 自分の読む目的と一致しているか考える。             |
| SRS  | I read slowly but carefully to be sure I understand what I'm reading.             | 自分が読んでいる内容を理解するために, ゆっくり, 注意深く読んでいる。    |
| *PSS | I discuss what I read with others to check my understanding.                      | 自分の理解を確認するため, 読んだ内容について, 他の人と話し合う。      |
| *GRS | I skim the text first by noting characteristics like length and organization.     | 文章の長さや構成を把握するために, まずひと通り流し読みをする。        |
| *PSS | I try to get back on track when I lose concentration.                             | 集中力がなくなったときは, なんとか立てなおそうとする。            |
| *SRS | I underline or circle information in the text to help me remember it.             | 英文に下線を引いたり, ○をつけたりして, その箇所を覚えておけるようにする。 |
| *PSS | I adjust my reading speed according to what I'm reading.                          | 読んでいる内容に応じて, 読むスピードを調整している。             |
| *GRS | I decide what to read closely and what to ignore.                                 | 英文を読むとき, 注意を向けて読む箇所とそうでない箇所を決める。        |
| SRS  | I use reference materials such as dictionaries to help me understand what I read. | 英文を理解しやすくなるように, 辞書のような参考資料を使う。          |

|      |   |   |
|------|---|---|
| PSS  | When text becomes difficult, I pay closer attention to what I'm reading.        | 英文が難しい場合は、読んでいる内容により注意を向ける。                 |
| *GRS | I use tables, figures, and pictures in text to increase my understanding.       | 表や図、絵などの情報を使用し、文(章)理解を深めている。                |
| *PSS | I stop from time to time and think about what I'm reading.                      | 時々読むのをやめて、何を読んでいるかについて考える。                  |
| *GRS | I use context clues to help me better understand what I'm reading.              | 文脈(文章の前後関係)からの手がかりを使用して、文(章)理解に役立てている。      |
| SRS  | I paraphrase (restate ideas in my own words) to better understand what I read.  | 文章の内容を言い換える(自分の言葉で表現すること)ことで、文(章)理解を促進している。 |
| PSS  | I try to picture or visualize information to help remember what I read.         | 読んだ内容を覚えておけるように、書かれている情景・状況を頭の中で思い浮かべる。     |
| *GRS | I use typographical aids like boldface and italics to identify key information. | 太字や斜体のような活字(文字)情報を利用し、要点を把握できるようにしている。      |
| *GRS | I critically analyze and evaluate the information presented in the text.        | 書かれている内容を、分析したり、評価したりする。                    |
| *SRS | I go back and forth in the text to find relationships among ideas in it.        | 読み返しをして、筆者の考えを読み取る。                         |
| *GRS | I check my understanding when I come across conflicting information.            | 自分の理解が文章の内容と矛盾しているときは、両者を確認する。              |
| *GRS | I try to guess what the material is about when I read.                          | 英文を読む際に、それがどんな内容なのか推測してみる。                  |
| PSS  | When text becomes difficult, I reread to increase my understanding.             | 英文が難しい場合、理解を深めるためにもう一度読み直す。                 |
| SRS  | I ask myself questions I like to have answered in the text.                     | 自問自答をしながら英文を読んでいる。                          |
| *GRS | I check to see if my guesses about the text are right or wrong.                 | 英文に対する推測が合っているかどうか確認する。                     |
| PSS  | I try to guess the meaning of unknown words or phrases.                         | わからない単語や表現(フレーズ)が出てきたら、その意味を推測しようとする。       |

(注) GRS = Global Reading Strategies; PSS = Problem Solving Strategies; SRS = Supportive Reading Strategies  
\*印がついた「観点」は、本研究の分析対象項目であることを示す。

### 3.3 眼球運動計測

眼球運動の測定には、SR Research 社製の EyeLink 1000を、刺激の呈示には、21インチの CRT ディスプレイ (EIZO FlexScan T965, 1024 × 768 pixel) を使用した(参照: 図1)。EyeLink 1000の視点測定サンプリング周波数は1000Hz, 1文字の視野角は画面上約0.33度に設定した。刺激呈示に際して、ディスプレイの背景色は灰色地に、文字色は黒文字に、フォントは Windows 標準の Arial 体に設定した。眼球運動測定時に参加者の頭部を固定するために顎台を使用し、参加者が顎台に顎をのせた状態で目からディスプレイまでの距離が65cmになるよう調整した(参照: 図2)。実験の制御および眼球運動データの記録には SR Research 社製の Experiment Builder を使用して実験プログラムを作成した。



▶ 図1: 本研究で使用した眼球運動計測装置(EyeLink 1000)とCRTディスプレイ



▶ 図 2：眼球運動計測時の実験参加者の様子

### 3.4 インタビュー調査

眼球運動測定後に半構造化インタビューを行い、読解力テスト解答時の認知プロセスについて内省を促した他、テスト項目に関する感想を述べてもらった。インタビューの内容は IC レコーダーに録音した。事前に設定した質問項目は以下のとおりである。しかし、参加者の回答によって、下記項目にはない点もさらに詳細に尋ねていく場合も頻繁にあった。

1. 空所補充（内容理解）問題を解答するときは、どのようなことに気を付けなければならないと考えていたか。
2. それは実行できたか。
3. 何を考えながら問題を解いたか。
4. 先に問題を見たか。テキストを見たか。
5. 何度も読み直さないとわからなかったか、一読で理解できたか。
6. テキストの内容を理解するのを重視したか、それとも問題を解くことだけに集中したか。
7. 空所補充（内容理解）の形式には慣れているか。
8. 全体的な難易度（7段階：1＝とても難しい、7＝とても簡単）
- 8.1 なぜそう感じるか。
9. 解答に対する自信（7段階：1＝全くない、7＝とてもある）
- 9.1 なぜそう感じるか。
10. 空所補充と内容理解の問題で、自分の中で読み方を変えたか。
11. その他感じたこと・感想・言いたいこと

## 3.5 手順

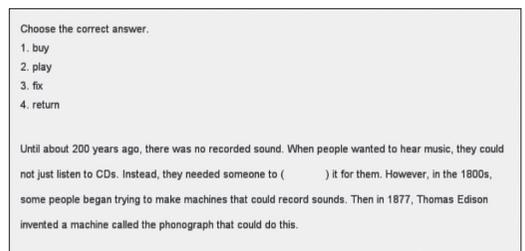
### 3.5.1 アンケート・インタビュー調査

アンケートは眼球運動計測後に単独での実施で、紙面上で解答を行った。所有時間は5分前後であった。インタビューは眼球運動計測直後に実験者（筆者）と参加者の1対1で行い、あらかじめ用意した質問項目について参加者が回答を行い、すべての項目に回答し終えるまで15分程度要した。上記の調査は眼球運動計測実験を実施した部屋と同じ防音室で行われた。

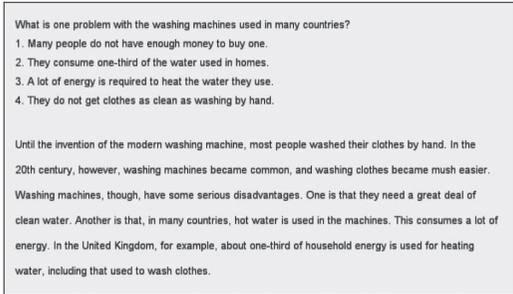
### 3.5.2 眼球運動計測実験

参加者は、防音室内に設置したディスプレイの前に着席後、ディスプレイの画面上に呈示される読解力テストに解答する指示を受けた。その後、眼球運動計測装置のキャリブレーション（精緻なデータを得るために、計測前に装置と参加者の眼球運動の調整を行うこと）を行い、以下の手順で実験を受けた。

1 試行は以下の流れで行われた。まず画面左上部に注視点として呈示された丸記号を注視すると、画面が切り替わり、画面上部に、空所補充問題の場合は指示文、選択肢4つの順に縦並びに、内容理解問題の場合は質問文、選択肢4つの順に縦並びに呈示され、画面中央から下部にかけてテキストが呈示された（参照：図3、図4）。その画面上で、参加者はテキスト読解を行い、コントローラーのキー押し、反応によって問題への解答を行うことが求められた。コントローラーの該当キーが押された後は、次の試行開始の注視点画面が呈示された。8試行を1ブロックとし、これを4ブロック行った。刺激項目の呈示はランダム順とした。本試行の前には、空所補充・内容理解問題の各1問を練習問題として解答し、テスト実施の手順を確認した。なお、実験実施上に関する質問はその時点ですべて解決された。読解テスト中の制限時間は設けず、参加者は各自の



▶ 図 3：空所補充問題のテスト画面



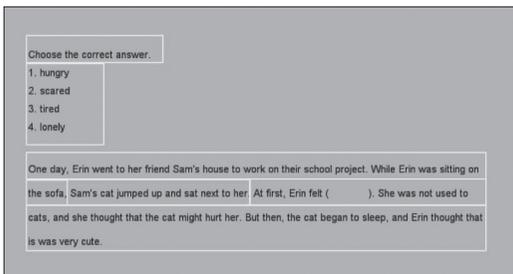
▶ 図 4：内容理解問題のテスト画面

ペースで解答を行った。32問全解答の所要時間は60分前後であった。

### 3.6 分析方法

読解テストの認知的妥当性を検証するため、Bax (2013) を参考に、正答者と誤答者のテストの解答中の眼球運動を比較する。眼球運動データの処理・分析は、以下の手順に従い行った。

- (1) 注視時間が80ミリ秒以下または1000ミリ秒以上である注視が除外された。
- (2) 解答する画面は、例えば、図5の空所補充問題のように、指示文 (Choose the correct answer), 選択肢 (1. hungry 2. scared 3. tired 4. lonely), テキストにおける解答該当箇所 (Sam's cat jumped up and sat next to her) と非解答該当箇所 (その他のテキスト) を、複数の関心領域 (Interest Area; 黄色の線で区切る (報告書は白黒印刷のため白線)) として分けられた。
- (3) 各関心領域における眼球運動の測定値は分析対象となる。
- (4) 眼球運動の測定値は、大きく分けて、注視時間、注視の回数、関心領域における注視時間・回数と1つの項目における合計の注視時間・回数 (総注視時間と総注視回数) の割合と、3種類に分けられた。本稿で報告する測定値は、以下



▶ 図 5：関心領域の設定仕様

の14個である；

1. 総注視時間；
2. 総注視回数；
3. 解答該当箇所：注視時間；
4. 解答該当箇所：注視回数；
5. 解答該当箇所：注視時間・総注視時間割合；
6. 解答該当箇所：注視回数・総注視回数割合；
7. 非解答該当箇所：注視時間；
8. 非解答該当箇所：注視回数；
9. 非解答該当箇所：注視時間・総注視時間割合；
10. 非解答該当箇所：注視回数・総注視回数割合；
11. 選択肢箇所：注視時間；
12. 選択肢箇所：注視回数；
13. 選択肢箇所：注視時間・総注視時間割合；
14. 選択肢箇所：注視回数・総注視回数割合。

Bax (2013) で使われていない「割合」に関する測定値を使用するのは、どの領域が比較的重要視されるかを知ることができるためである。

また、インタビューデータは、眼球運動データではとらえられない解答時の行為・行動・活動・気付きなどの内的な部分を描写しているため、眼球運動データの補完データとして使用した。

## 4 結果・考察

分析を行う前に、眼球運動計測時に実施上問題があった8名のデータを除外し、最終的に34名を分析対象とした。また、読解力テストにおいては、信頼性係数を著しく下げるテスト項目はあらかじめ分析対象から除外された。最終的分析対象となった項目数は20で、テスト得点平均値は17.79、標準偏差は2.28、信頼性係数 (クロンバックの  $\alpha$  係数) は.71であった。次節以降は、テスト得点からは明らかにできない読解力テスト解答時の認知プロセスを解明するために、眼球運動データの結果とその解釈を進めていく。

### 4.1 正答者と誤答者の眼球運動データの比較

まずリサーチクエスション1と2に取り組むために、Bax (2013) と同じように、正答者と誤答者の

テスト解答中の眼球運動データを項目ごとに比較した。しかし、表5のとおり、各項目の正答率が全体的に高かったため<sup>(注2)</sup>、1つの項目に対し誤答数が5つ以上の項目だけを絞って分析することにした。最終的に、比較対象となった項目数は10(項目番号5, 8, 9, 11, 14, 15, 23, 26, 28, 29)であった。

Bax (2013) と同様に、データ数の僅少の問題と、正規性の問題を考慮に入れ、分析は、ノンパラメトリックな統計学的検定——マン・ホイットニーのU検定——を使用した。10項目の中で、分析対象となる項目総注視時間、総注視回数、関心領域内における注視時間と注視回数において正答者と誤答者間で有意な差が見られたのは、項目5, 11, 15, 26であった。これら4項目が求める認知プロセスを表1に照合させると、表6のようにまとめられる。本稿はこれら4項目について詳細な結果の記述を行う。また、4項目ごとの眼球運動計測データの分析結果は表7から表10のとおりである。

■ 表5：項目合計統計量

| 項目番号 | 項目が削除された場合の尺度の平均値 | 項目が削除された場合の尺度の分散 | 修正済み項目合計相関 | 項目が削除された場合のクロンバック $\alpha$ |
|------|-------------------|------------------|------------|----------------------------|
| 1    | 16.79             | 5.20             | .00        | .71                        |
| 5    | 16.82             | 5.12             | .06        | .71                        |
| 7    | 16.94             | 4.78             | .18        | .71                        |
| 8    | 16.97             | 4.39             | .40        | .68                        |
| 9    | 16.88             | 4.83             | .22        | .70                        |
| 11   | 16.85             | 4.92             | .21        | .70                        |
| 13   | 16.79             | 5.20             | .00        | .71                        |
| 14   | 16.79             | 5.20             | .00        | .71                        |
| 15   | 16.88             | 4.89             | .17        | .71                        |
| 19   | 16.97             | 4.15             | .57        | .66                        |
| 20   | 16.91             | 4.63             | .33        | .69                        |
| 22   | 16.97             | 4.70             | .21        | .71                        |
| 23   | 16.79             | 5.20             | .00        | .71                        |
| 24   | 16.91             | 4.20             | .66        | .66                        |
| 26   | 17.38             | 4.24             | .34        | .70                        |
| 27   | 16.88             | 4.47             | .53        | .68                        |
| 28   | 16.94             | 4.36             | .47        | .68                        |
| 29   | 16.79             | 5.20             | .00        | .71                        |
| 30   | 16.85             | 4.98             | .15        | .71                        |
| 31   | 16.94             | 4.60             | .30        | .70                        |

■ 表6：最終分析項目の認知プロセスの階層レベル

|   | 認知プロセスの階層レベル | プロセスの説明                            | 対応する項目(問題の種類)   |
|---|--------------|------------------------------------|-----------------|
| 2 | 語彙：反意語の同定    | 選択肢とテキスト内にある反意語の特定                 | 5(空)            |
| 5 | 推論           | 文章間の情報を統合し、問題に正答するために必要な推論を行う      | 26(内)           |
| 6 | メンタルモデルの形成   | 文章間の情報を統合し、問題に正答するために必要な命題的知識を形成する | 11(空),<br>15(空) |

(注) (空) = 空所補充問題；(内) = 内容理解問題

#### 4.1.1 項目11への正答者・誤答者の眼球運動計測データの比較

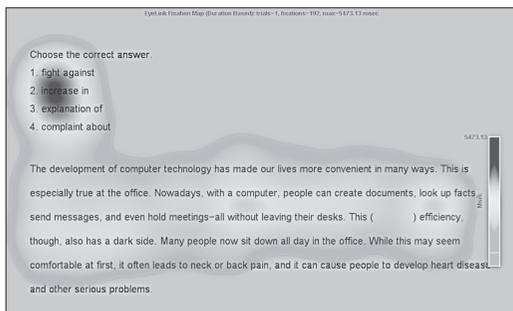
まず、メンタルモデルの形成を問う項目11(表7)については、誤答者は正答者よりも平均的に、総注視時間、解答該当箇所、非解答該当箇所および選択肢箇所における注視時間が有意に長く、総注視回数、非解答該当箇所および選択肢箇所に置かれる注視回数も有意に多かった。また、解答該当箇所における注視回数の違いについても誤答者と正答者で有意傾向が見られた。

図6と図7は項目11の誤答者Aと正答者Bの注視のヒートマップである。ヒートマップとは、解答時に注視が置かれた領域を可視化するためのツールの1つであり、より長く注視された領域から順に、赤→オレンジ→黄→緑の色順で表されている<sup>(注3)</sup>。図6と図7からわかるように、誤答者Aは、項目11の解答時に、より広い領域で注視を置いたのに対し、正答者Bは、より狭い領域に注視を置いていることから、正答者Bは、正答に導く情報のみに注意を払っていると見なすことができよう。また、誤答者Aが広範囲にわたって注視を置いたことから、正答を導く情報が特定できなかった、あるいは、文章間のつながりを理解して大きなレベルでのメンタルモデルを形成できなかったということが推測できよう。誤答者Aのインタビューデータからは、空所補充問題と内容理解問題の読み方の違いは、「穴埋め(空所補充)は途中で(読み)終わるけど、読解はとりあえずひと通り最後まで読むように」していたことであったが、共通の読み方は、最初に選択

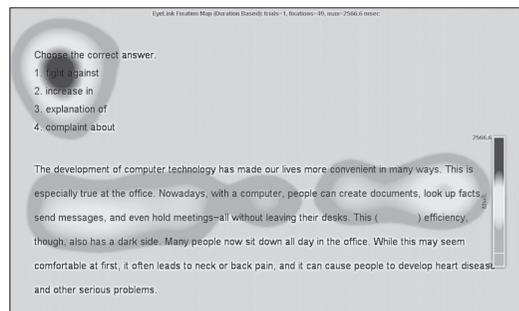
■表 7：項目 11 への誤答者と正答者の眼球運動計測データ

| 測定指標    | 誤答者 (n = 7)      |               | 正答者 (n = 27)  |               |        |        |             |         |
|---------|------------------|---------------|---------------|---------------|--------|--------|-------------|---------|
|         | 平均 (標準偏差)        | 中央値           | 平均 (標準偏差)     | 中央値           | U 値    | Z 値    | Sig. (両側検定) |         |
| 総注視時間   | 53619 (10960)    | 56347         | 37284 (16605) | 33057         | 41.000 | -2.280 | 0.023 *     |         |
| 総注視回数   | 202.0 (33.4)     | 204           | 145.9 (60.1)  | 138           | 41.000 | -2.280 | 0.023 *     |         |
| 解答該当箇所  | 注視時間             | 19322 (7496)  | 19056         | 12985 (5920)  | 12336  | 47.000 | -2.023      | 0.043 * |
|         | 注視回数             | 69.9 (20.7)   | 73.0          | 52.2 (21.1)   | 52     | 51.000 | -1.854      | 0.064 † |
|         | 注視時間・<br>総注視時間割合 | 0.351 (0.082) | 0.335         | 0.356 (0.075) | 0.345  | 91.000 | -0.149      | 0.881   |
|         | 注視回数・<br>総注視回数割合 | 0.345 (0.080) | 0.333         | 0.366 (0.072) | 0.353  | 81.000 | -0.575      | 0.565   |
| 非解答該当箇所 | 注視時間             | 19475 (4107)  | 21052         | 13809 (7718)  | 10806  | 42.000 | -2.236      | 0.025 * |
|         | 注視回数             | 76.9 (20.3)   | 77.0          | 51.5 (29.8)   | 43.0   | 41.000 | -2.281      | 0.023 * |
|         | 注視時間・<br>総注視時間割合 | 0.368 (0.078) | 0.360         | 0.358 (0.087) | 0.355  | 90.000 | -0.192      | 0.848   |
|         | 注視回数・<br>総注視回数割合 | 0.379 (0.073) | 0.368         | 0.338 (0.091) | 0.343  | 73.000 | -0.916      | 0.360   |
| 選択肢箇所   | 注視時間             | 14059 (3171)  | 13899         | 9270 (5222)   | 7675   | 38.000 | -2.406      | 0.016 * |
|         | 注視回数             | 51.7 (10.7)   | 51            | 36.4 (18.3)   | 28     | 41.000 | -2.281      | 0.023 * |
|         | 注視時間・<br>総注視時間割合 | 0.267 (0.062) | 0.274         | 0.248 (0.068) | 0.245  | 81.000 | -0.575      | 0.565   |
|         | 注視回数・<br>総注視回数割合 | 0.259 (0.056) | 0.260         | 0.250 (0.063) | 0.254  | 87.500 | -0.298      | 0.766   |

(注) \* $p < .05$ ; † $p < .10$ .



▶ 図 6：項目 11 誤答者 A の解答時のヒートマップ



▶ 図 7：項目 11 正答者 B の解答時のヒートマップ

肢を見て、次に本文を 1 行目から読み始めるということであった。また、誤答者 A の空所補充問題に対する難易度はやや「易しめ」であった。

一方、正答者 B のインタビューで明らかになったことは、正答者 B は「穴埋め問題（空所補充）の前後の文章をより深く理解しようとして」いたことで、図 7 のヒートマップのとおり、空所の前後のみに注視が置かれていることがわかる。また、正答者 B が空所補充問題に対して難しいとは全く感じず、自分の解答へ強い自信をも持っていた。その理由として「英文自体が短いのと、問題の選択肢にあ

る単語が簡単で、博打を打った（「勘で答える」という意味）ところなかった」と述べている。これらのインタビューデータから言えることは、誤答者 A は、空所補充問題に対してそれほど困難は感じていなかったが、正答者 B と比べて、読んでいるテキスト量が多いため情報量に統合させることが難しかったのではないかとということである。それに比べて正答者 B は、空所の前後に注意を払うことに意識を向けたため、余分な情報に惑わされることなく効率よく正答を導き出せた可能性がある。このような考察から言えることは、個人の読解時に用いる

ストラテジー（どこに注意を向けるか）の違いがテキストの総注視時間や正誤反応に影響している可能性があるということである。

#### 4.1.2 項目15への正答者・誤答者の眼球運動計測データの比較

項目15は、前述の項目11と同様にメンタルモデルの形成を問う項目であり、眼球運動計測データの結果は表8のとおりである。主な結果は、解答中の総注視回数および非解答該当箇所に置かれる注視回数の差が正答者・誤答者の間で有意傾向が見られ、誤答者が正答者より多くの注視を置く傾向にあるということである。

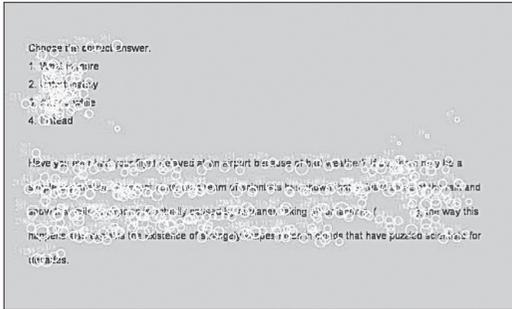
図8と図9は項目15に対する誤答者Cと正答者Dの解答時の注視パターンを示している。図中の丸の数は注視の回数を表している。表8の結果を裏づけるように、誤答者Cは正答者Dより、同じテキストに対し、全体的により多くの注視を置いていることがわかる。読解時の眼球運動に関する先行研究で報告されているように、単語の意味が理解困難で

ある場合、読み手は同じ単語に何回も注視する傾向にある（Rayner, 1998）。したがって、項目15の誤答者Cも、複数文のつながりを理解することが難しく、結果としてメンタルモデルを形成できず、正答にたどり着けられなかったのではないかと推測できる。しかしながら、誤答者Cの空所補充問題に関するインタビューデータでは、「後半になって疲れたりしたときは何回も読み直しはした」が、「話題とかも想像ができそうなやつだったので、良いことを言っているのか悪いことを言っているのか（肯定的な文章か否定的な文章か）で、だいたい（選択肢は）絞れたり、単語の意味とかが絶対に（解答として）ないなっていう感じで消去法とかもできたし、やりやすかった」と述べている。誤答者Cは、空所補充問題を解答する際は、空所の前までのテキスト内容をしっかり読み、解答がわかれば空所の後は読まないインタビューでは答えていたが、図8を見ると、テキスト全体に注視が置かれていることから、この項目に関しては、何らかの要因でその読み方がうまく機能しなかったことが示唆できる。

■ 表8：項目15への誤答者と正答者の眼球運動計測データ

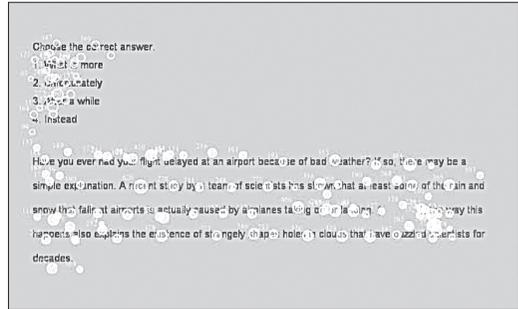
| 測定指標    | 誤答者 (n = 5)   |               | 正答者 (n = 29)  |               |        |        |                    |                    |
|---------|---------------|---------------|---------------|---------------|--------|--------|--------------------|--------------------|
|         | 平均 (標準偏差)     | 中央値           | 平均 (標準偏差)     | 中央値           | U値     | Z値     | Sig. (両側検定)        |                    |
| 総注視時間   | 68386 (22607) | 64954         | 50458 (20003) | 45932         | 41.000 | -1.532 | 0.126              |                    |
| 総注視回数   | 263.2 (75.8)  | 250.0         | 194.0 (79.3)  | 171.0         | 35.000 | -1.824 | 0.068 <sup>†</sup> |                    |
| 解答該当箇所  | 注視時間          | 7687 (3202)   | 6227          | 6104 (3383)   | 5943   | 51.000 | -1.045             | 0.296              |
|         | 注視回数          | 30.6 (13.5)   | 25.0          | 24.3 (13.7)   | 23.0   | 50.000 | -1.096             | 0.273              |
|         | 注視時間・総注視時間割合  | 0.114 (0.036) | 0.112         | 0.122 (0.053) | 0.114  | 64.000 | -0.413             | 0.679              |
|         | 注視回数・総注視回数割合  | 0.117 (0.040) | 0.109         | 0.125 (0.049) | 0.118  | 64.000 | -0.413             | 0.679              |
| 非解答該当箇所 | 注視時間          | 50902 (17388) | 43629         | 37021 (15312) | 33605  | 40.000 | -1.580             | 0.114              |
|         | 注視回数          | 192.2 (59.1)  | 160.0         | 138.1 (57.4)  | 131.0  | 37.500 | -1.702             | 0.089 <sup>†</sup> |
|         | 注視時間・総注視時間割合  | 0.745 (0.045) | 0.752         | 0.731 (0.060) | 0.734  | 65.000 | -0.365             | 0.715              |
|         | 注視回数・総注視回数割合  | 0.730 (0.061) | 0.731         | 0.711 (0.056) | 0.704  | 60.000 | -0.608             | 0.543              |
| 選択肢箇所   | 注視時間          | 9242 (4950)   | 7194          | 6268 (3505)   | 5175   | 42.000 | -1.483             | 0.138              |
|         | 注視回数          | 37.6 (20.3)   | 29.0          | 26.1 (14.6)   | 22     | 45.000 | -1.338             | 0.181              |
|         | 注視時間・総注視時間割合  | 0.134 (0.058) | 0.130         | 0.125 (0.041) | 0.117  | 72.000 | -0.024             | 0.981              |
|         | 注視回数・総注視回数割合  | 0.142 (0.069) | 0.144         | 0.134 (0.042) | 0.121  | 69.000 | -0.170             | 0.865              |

(注) <sup>†</sup>p < .10.



▶ 図 8：項目 15 誤答者 C の解答時の注視パターン

同様に、正答者 D も、「空所の前後を注意深く読んで」、「括弧のところまでで推測できるものはそこまで読むのをやめて、それでもわからなかったら最後まで読んでから」解答するストラテジーをとっていたが、正答者 D の場合は、このストラテジーが項目 15 への解答時にうまく作用していたと考えられる。正答者 D の選択肢を選ぶ際の内省時においても、「穴埋め（空所補充）は、（テキストを）読んで（選択肢を）さっと見て、これ、みたいな感じ」と答えていることから、項目 15 解答時も効率よく正



▶ 図 9：項目 15 正答者 D の解答時の注視パターン

答となる選択肢を選択していたと考えられる。

#### 4.1.3 項目 5 への正答者・誤答者の眼球運動計測データの比較

語彙レベルで反意語の同定を問う項目 5 の結果（参照：表 9）について、誤答者は正答者より、選択肢箇所において有意に長くかつ多く注視し、選択肢箇所における注視時間・総注視時間の割合と選択肢箇所における注視回数・総注視回数の割合が有意に大きかった。また、総注視時間に関しても、誤答者と正

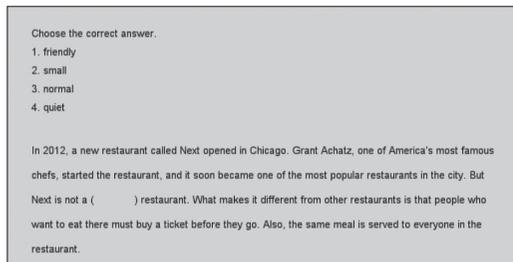
■ 表 9：項目 5 への誤答者と正答者の眼球運動計測データ

| 測定指標    | 誤答者 (n = 5)   |               | 正答者 (n = 29)  |               |        |        |             |         |
|---------|---------------|---------------|---------------|---------------|--------|--------|-------------|---------|
|         | 平均 (標準偏差)     | 中央値           | 平均 (標準偏差)     | 中央値           | U 値    | Z 値    | Sig. (両側検定) |         |
| 総注視時間   | 46570 (16176) | 50290         | 35540 (14097) | 35218         | 42.000 | -1.483 | 0.138       |         |
| 総注視回数   | 177.2 (51.5)  | 177           | 139.8 (50.5)  | 144           | 38.500 | -1.654 | 0.098†      |         |
| 解答該当箇所  | 注視時間          | 4935 (1827)   | 5090          | 4546 (3422)   | 3061   | 53.000 | -0.948      | 0.343   |
|         | 注視回数          | 20.6 (7.6)    | 22            | 15.3 (8.6)    | 12     | 43.500 | -1.414      | 0.157   |
|         | 注視時間・総注視時間割合  | 0.107 (0.032) | 0.096         | 0.125 (0.086) | 0.097  | 71.500 | -0.049      | 0.961   |
|         | 注視回数・総注視回数割合  | 0.117 (0.035) | 0.124         | 0.107 (0.039) | 0.101  | 53.000 | -0.948      | 0.343   |
| 非解答該当箇所 | 注視時間          | 32153 (9597)  | 35264         | 26323 (11615) | 24421  | 45.000 | -1.337      | 0.181   |
|         | 注視回数          | 122.4 (30.0)  | 119           | 104.4 (40.1)  | 110    | 48.500 | -1.168      | 0.243   |
|         | 注視時間・総注視時間割合  | 0.700 (0.051) | 0.701         | 0.739 (0.108) | 0.753  | 34.000 | -1.872      | 0.061†  |
|         | 注視回数・総注視回数割合  | 0.698 (0.050) | 0.689         | 0.745 (0.057) | 0.767  | 30.000 | -2.067      | 0.039*  |
| 選択肢箇所   | 注視時間          | 9197 (1604)   | 8567          | 3680 (1604)   | 3613   | 13.000 | -2.893      | 0.004** |
|         | 注視回数          | 32.8 (16.1)   | 33            | 14.8 (13)     | 13     | 14.500 | -2.829      | 0.005** |
|         | 注視時間・総注視時間割合  | 0.185 (0.050) | 0.170         | 0.109 (0.039) | 0.116  | 9.000  | -3.088      | 0.002** |
|         | 注視回数・総注視回数割合  | 0.177 (0.037) | 0.187         | 0.111 (0.038) | 0.109  | 14.000 | -2.845      | 0.004** |

(注) \*\* $p < .01$ ; \* $p < .05$ ; † $p < .10$ .

答者の間で有意傾向が見られた。

項目5は、空所補充問題（正答は3番の normal）であり、ここで問われているのは空所前後の not a ( ) restaurant と空所後の単語 different との関連、つまり not a の後に来る単語は different の反意語であるということを理解することである（参照：図10）。誤答者がより選択肢に注意を払った理由は、選択肢の中で different と対になる単語を特定するのに正答者よりも時間がかかったということが推測できる。



▶ 図10：項目5の問題

一方、項目5の非解答該当箇所における注視時間・総注視時間の割合については、正答者の方が誤答者よりも大きい有意傾向が見られた。これは正答者が非解答該当箇所をより長く注視していたと解釈するよりも、解答該当箇所を設定する際に適した箇所を選択していなかった実験実施上の問題があった可能性が高いと解釈する方がより妥当であると考えられる。というのも、項目が問う認知プロセスないし内容によっては、テキスト内の複数箇所が解答該当箇所に成り得るため、本研究実施時には選択しなかった箇所が、解答側にとっては、より解答を導くヒントとなる箇所であった可能性が考えられるためである。しかし、解答該当箇所をテキスト内の広範囲に設定してしまうと、本来調べたい認知プロセスを観察することは不可能になってしまう。そのため、このように意図していない結果が得られた理由は、解答該当箇所の選択方法によるものと考えの方がより妥当的だと思われる。

#### 4.1.4 項目26への正答者・誤答者の眼球運動計測データの比較

推論を問う項目26（結果は表10を参照）について、誤答者は、解答該当箇所における注視時間・総注視時間の割合が正答者よりも有意に低かったものの、

項目5と同じように、選択肢箇所には有意に長くかつ多く注視し、選択肢箇所における注視時間・総注視時間の割合と選択肢箇所における注視回数・総注視回数の割合が有意に大きかった。

## 4.2 メタ認知ストラテジーと眼球運動データとの関連

次にリサーチクエスチョン3に取り組むため、各項目に正答反応した眼球運動データを参与者ごとに平均化した値と、「読解ストラテジーに関するアンケート」（3.2.2節で詳述）で得られた参与者のメタ認知ストラテジーに関するデータとの相関関係に基づいて、読解力テストを解答する際の認知プロセスを考察する。ここで特に注目したい点は、参与者が習慣的に持つメタ認知ストラテジーと、正答に至るまでの解答プロセスとの関連である。なぜなら、読み手個人が所有するメタ認知ストラテジーの違いによって眼球運動パターンが異なる可能性が考えられるためである。Bax（2013）と本研究での正答者と誤答者の分析では、正答者と誤答者のテスト解答時の眼球運動の違いのみに着目したが、参与者の個人差に関する考慮が足りなかったことは否めない。そこで、参与者個人のメタ認知ストラテジーの使用とその種類が、解答時の眼球運動とどの程度関連しているかを調査するために、相関分析を行った。分析に使用したMARS（Mokhtari & Reichard, 2002）の3観点（Global Reading Strategies, Problem Solving Strategies, Supportive Reading Strategies）の記述統計は表11に示す。相関分析の結果は表12のとおりである。有意な相関関係が見られたのは、解答該当箇所の注視回数・総注視回数割合と Global Reading Strategies (GRS) であった ( $r=.41, p<.05$ )。GRSは、「テキストを読むときにどう考えるか」という観点であり、「英文を読むとき、注意を向けて読む箇所とそうでない箇所を決める」や「自分の理解が文章の内容と矛盾しているときは、両者を確認する」、「文脈（文章の前後関係）からの手がかりを使用して、文（章）理解に役立てている」などの質問項目から構成される。この観点と眼球運動計測データと有意な相関関係が見られたことは、上記の予測どおり、読み手が使用するメタ認知ストラテジーの種類と項目への正答解答プロセスとの間には関連があるということを意味している。この結果は、今後テスト項目の認知的妥当性を検証する上で

■ 表10：項目26への誤答者と正答者の眼球運動計測データ

| 測定指標    | 誤答者 (n = 6)   |               | 正答者 (n = 28)  |               |        |        |             |         |
|---------|---------------|---------------|---------------|---------------|--------|--------|-------------|---------|
|         | 平均 (標準偏差)     | 中央値           | 平均 (標準偏差)     | 中央値           | U値     | Z値     | Sig. (両側検定) |         |
| 総注視時間   | 77021 (29908) | 70325         | 57094 (21794) | 54591         | 49.000 | -1.581 | 0.114       |         |
| 総注視回数   | 281.7 (100.7) | 288           | 227 (81.6)    | 216.5         | 54.000 | -1.355 | 0.175       |         |
| 解答該当箇所  | 注視時間          | 11600 (2897)  | 11373         | 11519 (5713)  | 10973  | 67.000 | -0.768      | 0.442   |
|         | 注視回数          | 43.8 (7.73)   | 42.5          | 46.5 (21.3)   | 45.5   | 84.000 | 0.000       | 1.000   |
|         | 注視時間・総注視時間割合  | 0.160 (0.037) | 0.148         | 0.199 (0.047) | 0.192  | 39.000 | -2.033      | 0.042   |
|         | 注視回数・総注視回数割合  | 0.168 (0.048) | 0.160         | 0.204 (0.046) | 0.200  | 47.000 | -1.671      | 0.095*  |
| 非解答該当箇所 | 注視時間          | 28906 (9875)  | 28247         | 24941 (11181) | 23097  | 59.000 | -1.129      | 0.259   |
|         | 注視回数          | 98.7 (31.7)   | 104.5         | 90.4 (43.4)   | 77.5   | 65.500 | -0.836      | 0.413   |
|         | 注視時間・総注視時間割合  | 0.385 (0.065) | 0.367         | 0.438 (0.087) | 0.438  | 52.000 | -1.446      | 0.148   |
|         | 注視回数・総注視回数割合  | 0.357 (0.055) | 0.347         | 0.391 (0.076) | 0.389  | 52.000 | -1.446      | 0.148   |
| 選択肢箇所   | 注視時間          | 30267 (16588) | 28009         | 12988 (7384)  | 12519  | 28.800 | -2.530      | 0.009** |
|         | 注視回数          | 114.5 (57.8)  | 116.5         | 55.2 (27.9)   | 55.0   | 31.000 | -2.395      | 0.017*  |
|         | 注視時間・総注視時間割合  | 0.372 (0.089) | 0.409         | 0.225 (0.096) | 0.214  | 21.000 | -2.846      | 0.004** |
|         | 注視回数・総注視回数割合  | 0.387 (0.093) | 0.440         | 0.246 (0.097) | 0.235  | 21.000 | -2.846      | 0.004** |

(注) \*\* $p < .01$ ; \* $p < .05$ ; \* $p < .10$ .

■ 表11：MARSJの記述統計

|     | 項目数 | 平均 (標準偏差)   | $\alpha$ |
|-----|-----|-------------|----------|
| GRS | 13  | 3.27 (0.54) | .73      |
| PSS | 5   | 3.54 (0.59) | .69      |
| SRS | 8   | 2.66 (0.75) | .66      |

(注) GRS = Global Reading Strategies;  
 SRS = Supportive Reading Strategies;  
 PSS = Problem Solving Strategies;  
 $\alpha$  = クロンバックの  $\alpha$  係数

■ 表12：読解ストラテジーと眼球運動データ指標との相関 (n = 34)

| 測定指標    | SRS          | GRS  | PSS  |      |
|---------|--------------|------|------|------|
|         | 総注視時間        | -.11 | .16  | -.17 |
| 総注視回数   | -.05         | -.15 | .05  |      |
| 解答該当箇所  | 注視時間         | -.06 | -.06 | -.06 |
|         | 注視回数         | .04  | .27  | -.24 |
|         | 注視時間・総注視時間割合 | -.05 | -.15 | -.05 |
|         | 注視回数・総注視回数割合 | -.09 | .41* | -.32 |
| 非解答該当箇所 | 注視時間         | -.07 | -.14 | .07  |
|         | 注視回数         | -.07 | -.09 | .09  |
|         | 注視時間・総注視時間割合 | .01  | -.22 | .06  |
|         | 注視回数・総注視回数割合 | .02  | -.06 | .06  |
| 選択肢箇所   | 注視時間         | .07  | -.28 | .07  |
|         | 注視回数         | .09  | -.12 | .06  |
|         | 注視時間・総注視時間割合 | -.17 | -.16 | -.20 |
|         | 注視回数・総注視回数割合 | .07  | -.28 | .03  |

(注) \*\* $p < .01$ ; \* $p < .05$ .

は、個人差の要因をも同時に考慮する必要性を示唆している。

## 5 まとめ

本研究は、読解力テスト解答中の読み手（日本語を母語とし英語を学習する大学生および大学院生）の眼球運動を計測すると同時に、アンケート調査とインタビュー調査を行い、彼らのテスト解答時のテスト項目への認知的妥当性の検証と、認知プロセスの解明を試みた。本研究の主な結果は、

- (1) 項目正答者は、誤答者に比べて、解答時のテキスト注視時間が短く注視回数が少ないことから、解答該当箇所をより迅速にかつ的確に認識している；
- (2) テスト項目正答に至るまでの認知プロセスには読み手のメタ認知ストラテジーが関与している；
- (3) 解答時の認知プロセスを解明する上で、研究方法論的のトライアングレーションは有効に機能する、

であった。

(1)をさらに補足するならば、本研究では、Bax (2013) の実験実施上の問題点（参照：2.4節）を解決するために、英検の読解力問題のテキスト中の第1段落のみを分析対象としたが、その条件下においても、本研究が選定したテスト項目には、さまざまな認知プロセスの階層に属する項目が含まれていた。そのため、この点からも英検テスト項目には、解答者の能力を弁別するための最大限の試みがされていると言える。ゆえに、本研究を通して英検の読解問題項目の認知的妥当性が示されたと考えられる。

(3)については、3つの研究手法を使用したことで、それぞれのデータを補完することが可能であることを示唆できた。この点が本研究の強みである。例えば、「結果」の章で紹介した誤答者Cのインタビューデータでは、正答者Dと同様の内省が行われているが、その内容は、図8と図9のそれぞれの眼球運動データと一致はしていないと言える。ここから、テスト解答後にインタビュー調査のみを行って、そこから議論・考察を行うことは、当該参加者の読解プロセスを正しく反映していない可能性があるということが言える。この点は、質的調査を行う

研究者や教室指導を行う教員への具体的な提案につながると言えよう。

上記のように本研究を通して具体的な成果は得られたものの、いくつかの改善点が残される。まず、テストの難易度である。本研究は、準2級と2級のテキストを眼球運動計測の読解テスト問題として使用したが、表5の「修正済み項目合計相関」からわかるように、平均テスト得点が高く、参加者の能力を弁別できていない項目がいくつかある。これは、テスト項目の難易度が本研究の参加者にとって低かったことを意味する。インタビューデータからも、ほとんどの参加者が空所補充問題および内容理解問題に対し、「易しめ」だと回答していた。本研究における難易度選択は、本実験開始前に、顎台に顎を乗せて頭部を固定された状態で、計32題の読解力テストをパソコン上で解答をするという身体的な拘束が心身面の疲労を誘発するという可能性を考慮した結果であったが、結果的にその決断が本研究の参加者の能力を十分弁別できないこととなった。今後は、テストの難易度とテスト解答時の環境との関連をより検討する必要があると考えられる。しかしながら、難易度が低い試験級の項目を使用した場合にでも、図6と図7や、図8と図9、表7から表10で示したとおり、項目誤答者と正答者の眼球運動が有意に異なっている点は、当該項目が解答者に求める認知的妥当性が示されたことを同時に意味していると言える。

2つ目の問題点は、認知プロセスに与える要因についてである。本研究では、Bax (2013) に倣い、テスト解答時の認知的妥当性の検証を行ったが、今後の研究で求められることは、そのパフォーマンスに個人の能力（例：語彙力や文法力、ワーキングメモリ）がどのように関連しているのかを追求することである。

例えば、項目Nに対して同じ正答反応をした場合でも、テキスト総注視時間や総注視回数は個人間で多様に異なる。また、総合的な英語力を測定するテスト（例：TOEICやIELTS、TOEFL）では同程度のスコアであっても、解答に至るまでの内的なプロセスは複雑に異なると考えられる。つまり、眼球運動計測から浮き彫りとなった正答者と誤答者の認知プロセスの違いには、何らかの個人間の要因が深く関連しているということを意味している。読解プロセスおよび、そのプロセスをコントロールする個人

内の要因を解明することは、言語習得の観点から考慮すると、研究の必要性が高いと考えられる。

この点に関連するのは、アンケート調査で測定した3つの観点（GRS, SRS, PSS）と眼球運動計測データとの関係性についてである。本稿での報告は、眼球運動計測データと関連があった読み手の習慣的なストラテジーはGRSのみであった（参照：表12）。この結果の原因は、分析対象とした20項目すべての眼球運動計測データと各観点との相関関係を検討したことが考えられる。仮に、20項目すべてではなく、認知プロセスのレベルの階層化（Khalifa & Weir, 2009）に応じて項目を分類して、それぞれのレベルの項目とGRS, SRS, PSSの各観点との関連を調べると、異なる結果が得られる可能性がある。ゆえに、読解プロセスを解明するために本研究が今後取り組むべきことは、データをより詳細に分類し、再検討をすると同時に、個人の能力の側面を踏まえて検討することである。

## 6

### 結論

本研究は、第二言語習得研究のテストングの分野において眼球運動計測実験を取り入れると同時に

に、アンケート調査とインタビュー調査をも用いて、1つのデータを多角的に検討し、テスト解答時に認知プロセスの解明を試みた数少ない研究の1つである。本研究は、実験実施上における解決すべき問題点はあるものの、読解力テスト解答時の眼球運動計測の基礎的データを提供し、質的データと組み合わせることで、日本人英語学習者の読解認知プロセスを調査したことで、新たな視点からテスト評価と英語力評価の実現可能性を示唆した。

### 謝 辞

本研究遂行の機会を与えていただきました公益財団法人 日本英語検定協会の皆様と、選考委員の先生方、助言担当の吉田研作先生に心より感謝申し上げます。東京女学館大学の梁志鋭先生には、本研究の計画から実施、眼球運動データ分析時に親身なご指導をいただきました。また、名古屋大学の山下淳子先生には貴重な御助言や御意見を、加えて名古屋大学の杉浦正利先生には眼球運動計測機の使用許可をいただきました。この場をお借りして、改めて先生方に感謝申し上げます。最後に、調査実施に際しては、調査に協力していただいた42名の大学生、大学院生の皆様に厚く御礼申し上げます。

### 注

(1) Bax (2013) では、11項目の信頼性係数（クロンバックの  $\alpha$ ）は .722 と報告し (ibid. p.449)、実験で用いたテストは信頼性があると Bax はとらえているが、この数値は、眼球運動計測実験を行った参加者38名の IETLS テスト得点データに加えて、紙面上で同一のテストを実施した別の33名の得点データが加わったデータに基づいている。このような経緯に至った理由として考えられるのは、あくまで実験に使用したテストそのものの信頼性を調査したという考えである。しかしながら、Bax の研究で最終的な議論が行われているのは、眼球運動計測実験に参加した38名である。71名での信頼性係数と38

名での信頼性係数とは数値は異なるため（信頼性係数はサンプル数が多いほど高くなる傾向がある）、論文内で報告すべき信頼性係数は、71名分の得点データではなく、38名分の得点データに基づいて信頼性係数を出す必要があるのではないかととらえている。

(2) 表5内の「修正済み項目合計相関」の列で数値が .00 になっている場合は当該項目の正答率は100%であること、.0台の値を持つ項目の正答率は9割を超えていることを示す。

(3) 本紙のような白黒の場合、より長く注視された箇所ほど黒みが強く見られる。

## 参考文献 (\*は引用文献)

- \* Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30, pp.441-465.
- \* Bax, S., & Weir, C. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Cambridge Research Notes*, Cambridge ESOL, 47 (February 2012), pp.3-14. [http://www.cambridgeesol.org/rs\\_notes/rs\\_nts47.pdf](http://www.cambridgeesol.org/rs_notes/rs_nts47.pdf) より (最終閲覧: 2015年5月8日)
- \* Carrell, P. (1989). Metacognitive strategy training for ESL Reading. *TESOL Quarterly*, 23, pp.647-678.
- \* Dussias, P.E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, 30, pp.149-166.
- \* Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, 21, pp.175-198.
- \* Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.
- \* Khalifa, H., & Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.
- \* Kruger, J.L., & Steyn, F. (2013). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49, pp.105-120.
- \* Leung, C.Y., Sugiura, M., Abe, D. & Yoshikawa, L. (2014). The perceptual span in second language reading: An eye-tracking study using a gaze-contingent moving window paradigm. *Open Journal of Modern Linguistics*, 4, pp.585-594.
- \* Mokhtari, K., & Reichard, C.A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94, pp.249-259.
- \* Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, pp.372-422.
- \* Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, pp.1457-1506.
- \* Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, 28, pp.113-127.
- \* Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35, pp.213-235.
- \* Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17, pp.85-114.
- \* Weir, C., Hawkey, R., Green, T., & Devi, S. (2009). The cognitive processes underlying the academic reading construct as measured by IELTS. *British Council/IDP Australia IELTS Research Reports*, 9, pp.157-189.