

質問応答モデル QUEST に 基づく錯乱肢の作成

—解答収束メカニズムを利用して—

茨城県／筑波大学大学院在籍・日本学術振興会特別研究員 細田 雅也

申請時：茨城県／筑波大学大学院在籍

概要

人が質問に答えるとき、候補となる解答から最適なものを決める「解答収束メカニズム」が働く。QUEST はこの解答収束メカニズムによって、質問に対する最適解を生成する質問応答モデルである。本研究は QUEST による解答収束メカニズムを、英文読解テストにおける錯乱肢の作成に利用することを目的とした。調査 1 では、錯乱肢作成の基盤を築く目的で、QUEST の 3 つの解答収束メカニズム (i.e., リンク検索, 構造距離, 因果強度) がどのような順番で働くのかを検証した。調査 2 では、調査 1 から示された収束メカニズムの段階性 (解答収束ステージ) を利用して、収束メカニズム上最適解に近い、および遠い錯乱肢を作成し、テキスト理解度が異なる受験者の弁別にどの程度有効かを調べた。その結果、最適解に近い錯乱肢は、テキスト理解度の低い受験者をより多く引きつけることがわかった。これらの調査を通し、もっともらしい錯乱肢の性質に対する示唆と、質問応答モデルの言語テストへの利用可能性が示された。

1

はじめに

英文読解のテストにおいて、日々の授業から大規模テストに至るまで、最も多く使用されている形式の 1 つが、多肢選択式である。多肢選択式問題は、実行や採点が容易で、客観性を保証しやすいという利点がある。しかしながら、テキスト理解度の異なる受験者の弁別に有効な「もっともらしい (plausible)」錯乱肢 (不正解の選択肢) を作成することは決して容易ではない。この理由として、錯乱

肢の性質を実証的に検証した研究がこれまでのところ少なく、ゆえに、錯乱肢の有用性を客観的に規定する理論的知見が十分でなかった事情がある。

この背景を受けて本研究は、心理学の分野で開発された、質問応答モデル「QUEST」を、錯乱肢の作成に利用することに着目した。人が質問に答えるとき、さまざまな情報から最適な解答 (最適解) を決める「解答収束メカニズム (convergence mechanism)」が働く。QUEST は、この解答収束メカニズムを予測することで、最適解を生成するモデルである (e.g., Graesser & Clark, 1985; Graesser & Franklin, 1990)。実際に、先行研究は QUEST が人の質問応答を予測することを多く実証してきた (Graesser & Hemphill, 1991; Graesser, Lang, & Roberts, 1991; Millis & Barker, 1996)。しかしながら、その知見を言語テストに応用する試みはこれまでほとんどなかった。本研究は、QUEST に基づく錯乱肢を作成し、その有用性を検討することで、学習者の弁別に有効な錯乱肢の性質に示唆を与えることを目的とする。以下では、まず錯乱肢の性質に関する知見をレビューし、QUEST の概要を述べる。

2

先行研究

2.1 もっともらしい錯乱肢とは？

弁別力の高い錯乱肢は、受験者のパフォーマンスに対する精度の高い推測を可能にする。そのため、錯乱肢の作成は、多肢選択式問題の有効性を高める上で重要な役割を担っている。より具体的に、有効な錯乱肢とは、答えとしてもっともらしく、そのた

め理解の不十分な受験者がつつい正答として選択してしまうものといえる。実際に、Haladyna and Downing (1989)と Haladyna, Downing, and Rodriguez (2002) による多肢選択式問題のガイドラインによると、錯乱肢の作成における留意事項の1つに「もっともらしい錯乱肢であること」が提案されている。また、多肢選択式問題の難易度にかかわる要因を検討した Drum, Calfee, and Cook (1981)も、難易度を決める主要な要因の1つとして「錯乱肢のもっともらしさ (plausibility)」を指摘している。

このように、錯乱肢のもっともらしさの重要性は広く指摘されている一方、意外なことに、従来錯乱肢に対する定義はあいまいで、その構成要素を調べた例もあまりみられなかった。このことを受け、もっともらしい錯乱肢の具体的な性質を検証した研究に Ushiro et al. (2007) がある。ここでは、英文読解テストにおいて多くの受験者を引きつける錯乱肢がどのような要素から成り立っているかを、錯乱肢中における (a) 語数、(b) 内容語の割合、(c) テキストにない内容語の割合、(d) 否定語の数、(e) 前置された語の数、(f) 照応語の数の6つを考慮に検証した。実験では日本人大学生が TOEFL の読解問題に取り組み、50の多肢選択式問題に答えた。そして、上記の6要因を独立変数とした重回帰分析によって、錯乱肢の選択率がどの程度予測されるかを調べたが、いずれの要因も有意な予測力を得ることはできなかった。この結果から推測されることは、高木 (2011) が指摘するよう、語数などの表層的な性質は、錯乱肢のもっともらしさとはかわりが弱いということである。それでは、錯乱肢の有用性を決める性質とは、学習者の記憶や認知面とより深く関係するものであろうか？

高木 (2011) はこの点に対して、「記憶への保持されやすさ」という性質を観点に検証を行った。具体的には、テキスト中における他の情報との「因果的連続性」の強さが、錯乱肢の有用性を決める変数になると仮定された。因果的連続性とは、「A が起これば B も起こる」といえるとき、AB 間に想定される性質で、他との因果的連続性が高い情報はより強く記憶に残るとということが実証されている (e.g., Trabasso & Suh, 1993; Zwaan, Langston, & Graesser, 1995)。すなわち、因果的連続性が高い情報による錯乱肢は、解答時にテキストにあったと判

断されやすく、理解があいまいな受験者をより引きつけることが期待されるのである。高木は、英検の読解問題のテキストを読んだ日本人大学生を対象に、因果的連続性が高い錯乱肢、因果的連続性が低い錯乱肢、テキストにない情報による錯乱肢の3つを比較検証し、因果的連続性が高い錯乱肢は、テキストにない情報の錯乱肢よりも、テキスト理解度の低い受験者を多く引きつけることを示した。

以上の研究から得られた示唆をまとめると、読解テストにおけるもっともらしい錯乱肢の条件として、(a) テキスト中の情報を使用していること、(b) 受験者の記憶と認知的に関連していることの2つを仮定することができる。本研究ではこの両方を同時に満たす点で、質問応答モデル QUEST の利用可能性に注目した。続く節では QUEST の概要を、本研究と関連する部分に焦点を当てて述べる。

2.2 質問応答モデル QUEST

質問応答モデルとは、特定の質問に対する解答を、理論的に想定された原理に基づいて予測するものである。QUEST は1980年代から開発、発展されてきた質問応答モデルで、物語文 (Graesser et al., 1991)、説明文 (Graesser & Hemphill, 1991; Millis & Barker, 1996)、インタビュー (Graesser, Roberts, & Hackett-Renner, 1990) などにおける質問応答を有効に予測することが実証されている。そのため、QUEST は質問紙の作成や、コンピュータチューターの開発へ理論的基盤を与えてきた (e.g., Graesser, VanLehn, Rosé, Jordan, & Harter, 2001)。その一方、質問応答モデルの知見をテスト項目の作成に応用する試みは、これまでのところあまりなく、本研究ではこの可能性に取り組むことを目的とする。

QUEST は非常に複雑な理論要素から構成され、そのすべてを扱うことは本研究の範囲を逸脱するため、ここでは本研究と関連するものにのみ焦点を当てて概説することとする。例として、無重力空間が人体に与える影響を記述したテキストの読み手に対して「なぜ宇宙飛行士は喉が渇かないのか」という質問が与えられた場合を考えてみたい。QUEST によってこの質問への答えが生成される過程を大まかに述べると、まずこの質問文のカテゴリーが特定される (i.e., Why-Event 質問)。次に、質問が属する情報源 (i.e., 読んだテキストの記憶) において、解

答の候補（以下、解候補と述べる）が、特定された質問カテゴリーに基づいて活性化される。そして、活性化された解候補は、解答収束メカニズムによって少数の最適解（e.g., 体が水分量を減らそうとするから）に絞られるという過程を経る。以下では、QUEST が解答を産出する仕組みをより具体的に知るために、これらの構成要素（i.e., 情報源、質問カテゴリー、解答収束メカニズム）の概要を紹介する。

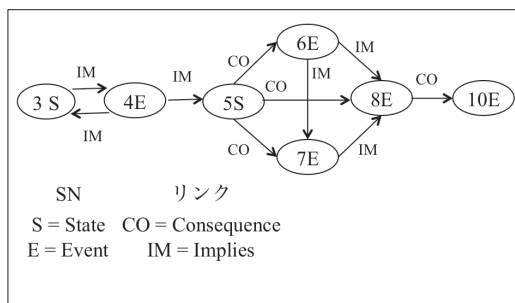
2.2.1 情報源：概念図構造

QUEST には解答がどのような知識構造から出てくるかを想定するために、質問と解答が属する「情報源（information sources）」が仮定されている。より具体的に、テキストに対する質問と解答の情報源は、読み手により構築されたテキストに対する記憶を指す（Graesser & Hemphil, 1991）。そして、この情報源は、「概念図構造（conceptual graph structures：CGS）」と呼ばれる図として操作化される。本研究で用いたテキストの一部を例に CGS の具体例をみていきたい（表1、図1を参照）。ここでは便宜上、日本語のテキストで考えることとする。

CGS は、基本的な情報単位である「ステートメントノード（statement nodes：SN）」が、情報間の関係性を示すリンクによって相互につながる構

■ 表1：ステートメントノード（SN）の例

カテゴリー	No.	SN
State	3	人体は複雑である。
Event	4	人体は無重力に自動的に反応する。
State	5	宇宙では重力は人体に影響を与えない。
Event	6	血や水分は下半身に巡らない。
Event	7	血や水分は足に巡らない。
Event	8	血や水分は上半身に移動する。
Event	10	胸や頭が血や水分で満たされていると体を感じる。



▶ 図1：SN3からSN10までの関係性を示すCGS

造を持つ。まずSNは図1のCGSにおいて円で示されるものである。これは1つの述部と複数の項から構成される情報単位で、それぞれに種類が付けられている。本研究にかかわるのはState, Event, Actionの3種類である。まず、Stateは与えられた時間軸において変化のない状態（e.g., 宇宙には重力がない）を意味する。対してEventは状態変化を表し（e.g., 心臓が小さくなる）、Actionは何らかの意図を伴う行動を意味する（e.g., 宇宙飛行士はあまり水を飲まない）。

次に、図1でSNをつないでいる矢印に注目されたい。これは、SN間の関係性を示すリンクで、それぞれに種類と方向が付けられている。本研究ではConsequence, Implies, Reason, Outcome, Initiateの5種類を扱う。まず、Consequence（CO）は原因と結果の関係にある情報同士をつなぐ役割を持つ（e.g., 宇宙では重力が人体に影響しない—CO→血や水分が下半身に巡らない）。次に、Implies（IM）は意味が包含関係にあるSN同士をつなぐ（e.g., 血や水分は下半身に巡らない—IM→血や水分は足に巡らない）。より具体的に述べると、因果関係を示すCOでは前のSNが後ろより時系列的に前に起こっている必要があるが、IMでは共起関係が示され、同じタイミングで発生するSN同士がつながれる。Reason（R）は、上下関係にある2つの目標行動をつなぎ（e.g., 肺や心臓が指令を出す—R→体が上半身の水分量を減らす）、Outcome（O）は行動とそれにより生じた事象をつなぐ（e.g., 上半身の水分量を減らす—O→水分が排除される）。そして、Initiate（IN）は事象とその結果起こった行動とをつなぐ（e.g., 宇宙飛行士は喉が渇かない—IN→宇宙飛行士はあまり水を飲まない）。このようなSNとリンクからなるCGSは、QUESTが質問に対する解答を生成する基盤として位置づけられる。以下ではそれが機能する仕組みをみていきたい。

2.2.2 質問カテゴリー

QUESTは、与えられた質問をCGS上のSNと対応づけることによって質問のカテゴリーを決定し、解答に妥当な情報を特定する。具体的には、質問カテゴリーは質問の基幹部（e.g., why, how, when）と、質問対象となるSN（以下、質問ノードと述べる）の種類（i.e., Event, Action）の組み合わせによって規定される。例として「なぜ宇宙で体は胸や頭が血

や水分で満たされていると感じるのか」という質問を考えてみたい。表1をみると、この質問に対応するSNはEVENTのSN10（胸や頭が血や水分で満たされていると体を感じる）である。そして、これをWhyによって尋ねていることから、Why-Event質問と分類される。ここで大切なことは、このように質問カテゴリを特定できれば、その解答として理論的に妥当な情報を、CGS上において示唆できるという点である。これを可能にするのが以下に述べる解答収束メカニズムである。

2.2.3 解答収束メカニズム

(1) リンク検索手続き

上記のとおり、QUESTによると「なぜ宇宙で体は胸や頭が血や水分で満たされていると感じるのか」という質問はWhy-Event質問と分類される。そして、Why-Event質問とは一般的に述べると、質問対象のイベントが起こった原因を尋ねるものである。このことから、このカテゴリの質問に対する妥当な解答とは、質問ノード（i.e., SN10）の原因に当たる情報と考えられる。では、SN10の原因に当たる情報はどれであろうか。ここで利用するのが、SN同士の関係性を示すリンクである。具体的には、QUESTによるとSN10の原因に当たる情報は、SN10からCO, IM, O, INの種類のリンクを後ろ向きにたどる途上にあると仮定されている。これには、SN10以外のすべてのSNが位置しており、よってこれらが上記のWhy-Event質問に対する「妥当解（legal answer）」として考えられる。

このような、質問カテゴリに基づく解答の検索手続きを、表2にまとめる。上で述べたとおり、QUESTは、特定の種類のリンクを特定の方向（この一連のリンクをlegal pathと呼ぶ）にたどることによって、各質問カテゴリに対する妥当解を生成する。これが、QUESTが想定する3つの解答収束

■ 表 2：質問カテゴリとその性質

質問カテゴリ	解答として妥当な情報	legal path
Why-Event	原因情報	←CO, ←IM, ←O, ←IN
Why-Action	目標情報	→R, →O
	原因情報	←CO, ←IM, ←O, ←IN

(注) CO = Consequence, IM = Implies, R = Reason, O = Outcome, IN = Initiate

メカニズムの1つ「リンク検索手続き（arc-search procedures）」である。以下では簡潔に、リンク検索と述べることにする。

端的に表すと、リンク検索は解答が質問に対し論理的に正しい関係にあることを規定している。例えば、原因を尋ねるWhy質問に対して、その原因や理由の情報が、結果や結論の情報よりも答えとして適切（i.e., 妥当解）になると想定しているのである。この性質のため、リンク検索はQUESTの解答収束メカニズムの中でも、根本的なものとして位置づけられる。

実際に、リンク検索を満たす妥当解の「解答としての良さ（Goodness of Answer : GOA）」は、リンク検索を満たさない「非妥当解（illegal answer）」よりも高く評価されることが、先行研究において広く実証されている（Graesser & Hemphill, 1991; Graesser et al., 1990; Graesser et al., 1991）。特に、説明文に対する質問応答を扱ったMillis and Barker (1996)は、リンク検索の想定（i.e., 妥当解が非妥当解よりも良い解答と評価される）が、読み手による評価と一致することを、QUESTがパフォーマンスに適合する条件として扱っており、このメカニズムが解答収束の根幹にあることを支持している。

(2) 構造距離

上記のとおり、リンク検索は中心的な解答収束メカニズムであるが、これを満たす妥当解の中でも、解答としての適切さには程度差があると考えられる。例えば、原因を尋ねるWhy質問（e.g., なぜ宇宙で体は胸や頭が血や水分で満たされていると感じるのか）に対し、より直接的な近因の答え（e.g., SN8 血や水分が上半身に移動するから）は、遠因の答え（e.g., SN3人体は複雑だから）と比べて、より適切な解答といえるのではないだろうか。このような、質問対象からの距離に基づく収束メカニズムは「構造距離（structural distance）」と呼ばれる。一般的に述べると、GOAは質問ノードからの距離に反比例し、解答が質問ノードに近ければ近いほど、適切な解答と評価される傾向にあるといえる。実際に、多くの先行研究は質問ノードからの距離とGOAとの間に負の相関を報告している（e.g., Graesser & Clark, 1985; Graesser & Hemphill, 1991; Graesser et al., 1990; Graesser et al., 1991）。

(3) 因果強度

以上のリンク検索と構造距離に加え、QUESTには「制約充足 (constraint satisfaction)」という収束メカニズムが仮定されている。これは、質問と解答との意味的、概念的な適合度をいくつかの観点から規定するものであるが、本研究ではそのうち「因果強度 (causal strength)」に焦点を当てる。この理由は、テキスト理解研究において因果関係の重要性は広く支持されており (e.g., McCrudden, Schraw, & Lehman, 2009; Trabasso & Suh, 1993), また高木 (2011) において有用な錯乱肢の性質として因果的連続性が実証されているからである^(注1)。このメカニズムによると、質問解答間の因果強度に従って、GOAは高くなる (Graesser et al., 1991)。ただし、因果強度の影響は、テキストや読み手の要因に左右され得ることが先行研究から示唆されている。例えば、物語文を用いた Graesser et al. (1991) では、因果強度は GOA を有効に予測したが、科学的な内容の説明文を材料とした Graesser and Hemphill (1991) では、因果強度による影響は確認されなかった。この点について Graesser and Hemphill は、因果強度が影響力を持つには、質問解答間の因果関係に対する正確な理解が求められるためと考察している。すなわち、なじみのない内容の説明文に対し因果的に精緻な理解を構築することは、物語文と比べて難しく、因果強度の影響が現れにくい場合があるといえる。

2.2.4 収束メカニズムの順序性と解答収束ステージ

前節で概観した3つの解答収束メカニズム (i.e., リンク検索, 構造距離, 因果強度) を通し, QUEST は特定の質問に対する最適解を生成する。しかしながら, これら3つが働く上で, 何らかの優先順位は存在するのであろうか。例えば, リンク検索が構造距離より先に解答を絞ると仮定すると, 構造距離はリンク検索にパスした妥当解をさらに収束させるために機能すると考えられる。そのため, 構造距離の影響は, 非妥当解よりも, 妥当解において顕著になると推測される。このような収束メカニズムの順序性を検証した希少な研究が Graesser et al. (1991) である。ここでは, 平易な物語文を読解した英語母語話者が, 物語の内容に関する質問と解答のペアを複数与えられ, 質問に対する各解答の良

さ (i.e., GOA) を評価した。そして, 重回帰分析によって, GOA に対するリンク検索, 構造距離, 制約充足の3つの収束メカニズムの予測力が調べられた。その結果, 収束メカニズム間の交互作用から Why 質問に関して, リンク検索→制約充足→構造距離の順番でメカニズムが働くことが確認された^(注2)。この結果に基づくと, Graesser et al. (1991) が検証した Why 質問には, 表3のような解答収束メカニズムによる段階, すなわち「解答収束ステージ」を仮定できる。

■表3: Graesser et al. (1991) の Why 質問における解答収束ステージ

ステージ1	リンク検索: 非妥当解から妥当解を区別
ステージ2	制約充足: 妥当解のうち, 質問と意味的, 概念的に一致度が高いものへ収束
ステージ3	構造距離: ステージ2をパスした解候補のうち, 質問ノードに近いものへ収束→最適解

ここで重要な点は, この解答収束ステージを利用することで, 最適解への距離が操作された解答を作成できるということである。例えば, リンク検索 (ステージ1) と制約充足 (ステージ2) はパスするが, ステージ3の構造距離が遠い解答は, 収束メカニズム上において最適解により近いものといえる。一方, ステージ1のリンク検索に違反する解答は, 最適解から遠いものといえるであろう。そして, 解答収束メカニズムが働く基盤となるのは, 読んだテキストに対する記憶, すなわち, 読み手によるテキスト理解である。テキスト理解がテキスト内容に対し一貫していれば, CGS のようなテキスト情報による精緻なネットワークを心内に表象できると考えられるが, テキスト理解がテキスト内容に対しあいまいな受験者は, そのようなネットワークを構築できていない可能性がある。後者の場合, 受験者はあいまいな理解に基づいて, 最適解とは別の解答へ誤って収束してしまうのではないだろうか。

この可能性は興味深い, 収束メカニズムの順序性を詳細に調べた研究は, Graesser et al. (1991) 以外ほとんどみられないことに注意する必要がある。すなわち, なじみのある内容で, 構造的にも平易な物語文ではこのような順序性が確認されたが, 英文読解テストでしばしば用いられる複雑な説明文に対して, 同様のことがいえるとは限らない。本研

究は、次節に述べる観点からこの点に取り組み、QUEST を利用した錯乱肢の有用性を明らかにする。

3

本研究の概要と目的

ここまで、先行研究から示唆される有用な錯乱肢の性質と、QUEST が解答を収束させるまでのメカニズムをみてきた。本研究の中心となるアイディアは、3つの解答収束メカニズム（i.e., リンク検索、構造距離、因果強度）が働く順番を考慮することで、最適解に近い、および遠い解答を作成し得るという可能性である。そして、収束メカニズムの基盤となるテキスト理解が不十分な受験者は、情報間の関係性の理解があいまいなため、そのような解答に誤って収束してしまう可能性が考えられる。これらの可能性に取り組む目的で、本研究は2つの調査を行った。まず調査1では、錯乱肢作成の前段階として、3つの収束メカニズムがどのような順番で働くかを検証した。続く調査2では、調査1から示された最適解に収束メカニズム上近い、および遠い解答が、錯乱肢として有用か否かを検証した。

4

調査 1

4.1 目的

調査1の目的は、英語学習者の質問応答において3つの解答収束メカニズム（i.e., リンク検索、構造距離、因果強度）がどのような順番で働くかを明らかにすることである。具体的には、Graesser et al. (1991) で用いられたものよりも長く複雑な説明文を材料として用い、その内容に対する質問応答を検証する。

4.2 方法

4.2.1 協力者

国立大学に通う55名の日本人大学生・大学院生が調査に参加した。協力者の専攻は、生物学、心理学、人文学、医学、社会学、教育学など多岐にわたる。

4.2.2 マテリアル

(1) テキスト

説明文読解に関する先行研究から、無重力が人体に与える影響を記述した説明文を採用した（e.g., McCrudden et al., 2009; Ushiro et al., 2015）。このテキストを選んだ理由は、QUEST の概念図を作成するには、適度に明確な構造を持つものが望ましいからである（Graesser & Clark, 1985）。

ここで注意すべきは、テキスト内容に対する質問への答えが収束していく過程を正確に検討するには、その根底にあるテキスト理解が確かであることが求められる点である。仮にテキスト理解が不正確なら、先に述べたとおり、解答は不正確な理解に基づいて内容的に誤ったものへ収束し、結果として観察された収束メカニズムは歪んだものとなってしまいう可能性が考えられる。このことから、調査1ではテキストを協力者の母語である日本語に訳して用いた（表4を参照）。字数は439字であった。

■ 表4：調査1で用いたテキスト

人が初めて宇宙旅行を考えたとき、宇宙の無重力が人体にどのような影響を与えるかはわからなかった。人体は複雑で、無重力に自動的に反応する。宇宙にいと、人体は重力の影響を受けない。ゆえに、血や水分は、下半身、特に足に巡らなくなる。実際に、体内の水分は上半身に移動する。このことは興味深い現象を起こす。血や水分が上半身に移動するため、体は、胸や頭が血や水分で満たされていると感じる。このため、肺や心臓は、上半身の血や水分を減らすようメッセージを出す。その結果、宇宙飛行士は喉が渇かなくなる。したがって、宇宙飛行士はあまり水を飲まない。体内の水分が排出されるため、体の水分レベルは通常より低くなる。血や水分の量が減ると、人体は通常どおりに働くのが難しくなる。したがって、心臓は地上にいるときほど強く働かなくなる。その結果、心臓は小さくなる。

宇宙の研究の成功は、人体が宇宙環境からどのような影響を受けるかを解明することにかかっている。宇宙旅行が人体に与える影響を研究する別の利点は、高血圧や心臓病など、地上の人々が苦しむ多くの病気をよりよく理解することに役立つことである。

(2) CGS

この説明文に対するCGSを、Millis and Barker (1996) の手法に従って作成した。なお、この段階ではテキストは英語の状態で作業を行った。

まず、調査者がテキストを23のSNに分割した

(資料1；詳しい手順と基準については Graesser & Clark, 1985を参照)。先行研究による基準に加え、本調査ではテスト項目として利用するために、質問や解答として十分なまとまりを持つことを条件とした。続いて、調査者と英語教育学を専攻する大学院生の2名によって、SNをState, Event, Actionにそれぞれ6, 13, 4ずつ分類した。評価者間一貫性は $Kappa=0.94$ で^(注3)、不一致点は協議により解決した。そして、種類が特定されたSNを調査者と英語教育学を専攻する大学院生の2名が、テキストを参考にしながらリンクでつなぎ、CGSの試案を別々に作成した。作成されたCGSの信頼性を検討するために、Millis and Barker (1996) に従い、2名が同じ種類かつ同じ方向のリンクでつないだノードの割合を計算したところ68%の一致率が得られた。リンクとSNの組み合わせには膨大なバリエーションがあることを考えると、これは十分な値といえよう(Millis & Barker, 1996, p.66を参照)。不一致点は協議により解決し、最終的なCGSを決定した。CGSの全体像は資料1を参考にされたい。

(3) GOA 評定課題：質問解答ペア

解答収束メカニズムを調べるために、テキスト内容の質問に対する解答の良さを1(悪い)から5(良い)で評価するGOA評定課題を冊子にて作成した。以下ではまず、質問と解答の作成手順を述べる。

理論上、QUESTはWhy, How, Whenなど多様な質問カテゴリーを検証できるが、本調査では理解と最も密接にかかわっており、かつモデルと実際の質問応答との適合度が最も高いと実証されているWhy質問に焦点を当てた(e.g., Graesser & Clark, 1985; Graesser & Hemphill, 1991; Millis & Barker, 1996)。QUESTが質問対象として扱うのは、EventとActionの2種類に属するSNである(Graesser et al., 1991; Millis & Barker, 1996)。CGSの中で17のSNがこの2種類に属していたが、意味的に重複したものを除いた14を質問対象、すなわち質問ノードとして選んだ。テキストの性質上そのほとんどがEventであった(Event: $n=12$, Action: $n=2$)。

質問解答ペアの作成においては、まず各質問ノードにWhyをつけてWhy質問文を作成した(e.g., The heart becomes smaller. → Why does the heart become smaller?).次に解答を、質問周辺のSNに基づき作成した。具体的には、Graesser and

Hemphill (1991) に従い、質問ノードがEventの場合、原因を示すBecauseを解答に用いるSNの文頭に加えた(e.g., The heart does not work as hard as on Earth. → Because the heart does not work as hard as on Earth.).一方、質問ノードがActionの場合、Becauseに加え、行動目的を示すIn order toを伴う解答を作成した(e.g., Some organs try to reduce the amount of blood and water in the upper body. → In order to reduce the amount of blood and water in the upper body^(注4))。そして、質問の理解に支障が出ないように、質問解答ペアはすべて日本語に訳された(Becauseは「～から」、In order toは「～するために」とした)。調査者と英語教育学を専攻する大学院生の2名で訳出を行い、不一致が生じた箇所は協議により解決した。その結果、合計120の質問解答ペアが作られた。SNが持つ原因や結果の情報の数はそれぞれ異なるため、1つの質問に対する解答の数には5から12のバリエーションがあった($M=8.57$, $SD=1.95$)。

次に、これらの解答に対して、3つの収束メカニズム(i.e., リンク検索, 構造距離, 因果強度)にかかわる性質を特定した。まず、リンク検索について、質問ノードからQUESTが妥当と想定する種類と方向のリンク上(i.e., legal path)にある解答は妥当解(リンク=1)、反対に、legal pathの逆方向のリンク上にある解答は非妥当解(リンク=0)とした(妥当解: $n=68$, 非妥当解: $n=52$)。例えば、「なぜ宇宙で血や水分は上半身に移動するのか」という質問を考えてみたい。これはSN8に対するWhy-Event質問で(表1を参照)、legal pathはCOとIMのリンクを逆行するものである(表2を参照)。よって妥当解には、この上にあるSN3から7が該当する。反対にSN10は、legal pathと反対方向のリンク上にあることから、非妥当解となる。次に構造距離について、調査1では質問ノードとの隔たりが2リンク以内の解答を構造・近(距離=0)、4リンク以上の解答を構造・遠(距離=1)とした(構造・近: $n=67$, 構造・遠: $n=53$)。図1のSN10を例に挙げると、2リンク以内にあるSN5, 6, 7, 8が構造・近、4リンク分遠くにあるSN3が構造・遠となる。最後の因果強度は、Graesser and Hemphill (1991) に従い、van den Broek, Trabassoらの基準によって計算され、0から1の値をとった(Trabasso & van den Broek, 1985; van den Broek,

1990)。調査者と英語教育学を専攻する大学院生の2名が各質問解答ペアの因果強度を算出し^(注5)、不一致点は協議により解決した($M=0.46, SD=0.28$)。そして、これらの質問解答ペアはGOA評定の記入欄とあわせて冊子に印刷された。

4.2.3 手順

調査は個別に実施され、ほとんどの場合30分以内に終了した。協力者は、内容を理解するよう注意しながらテキストを2度読解し(Millis & Barker, 1996)、GOA評定課題に取り組んだ。1つの質問に対する解答の提示順はランダムで、各解答が与えられた質問に対してどのくらい「良い答え」になっているかを1(悪い)から5(良い)で評価するよう指示された。また、評定の基準を明確にする目的で、「良い答え」とは「文章の文脈から考えて正しく、質問が尋ねていることに対し適切な情報を与えているもの」、反対に「悪い答え」とは「文章の文脈から考えて間違っており、質問が尋ねていることに対し不適切な情報を与えているもの」という定義が与えられた(Graesser & Hemphill, 1991)。さらに、評定はテキストの内容に基づいて行うよう指示された。

4.2.4 採点・分析

まず、質問に対する各解答の良さを知るために、各質問解答ペアのGOA評定値の平均を算出した。以下、これをGOAスコアと述べる。そして、このGOAスコアに3つの解答収束メカニズム(i.e., リンク検索、構造距離、因果強度)が、それぞれどの程度貢献したかを、GOAスコアを従属変数とした階層的重回帰分析により調べた。この分析では、独立変数を複数の段階(ステップと呼ばれる)に分けて投入することで、特定の変数による効果に焦点を当てた検証が可能となる。本調査では、ステップ1で3つの収束メカニズム(リンク検索: 妥当解=1, 非妥当解=0; 構造距離: 近=0, 遠=1; 因果強度=0~1)、続くステップ2で、これらの収束メカニズムを掛け合わせた4つの交互作用項(i.e., リンク×距離、リンク×因果、距離×因果、リンク×距離×因果)を投入した。これによって、各収束メカニズム自体による効果を取り除いた後の交互作用に注目した検証を行った(Graesser et al., 1991)。

本研究の統計分析では、有意水準を一貫して $\alpha=.05$ に定めた。また、3水準以上の比較では、第

一種の過誤を避けるために、ボンフェローニの方法によって有意水準を調整した。

4.3 結果と考察

リンク検索と構造距離に基づくGOAスコアの記述統計を表5に、GOAスコアと収束メカニズム、およびその交互作用との相関関係を表6に示す。階層的重回帰分析の結果、ステップ1において3つの収束メカニズムはGOAスコアの61.8%を説明しており($p<.001$)、ステップ2で追加された交互作用項は、説明率を4.3%向上させた($p=.008$)。各収束メカニズムと交互作用の説明率に関しては、リンク($\beta=.66, p<.001$)、距離($\beta=-.36, p=.030$)、そして、リンク×距離の交互作用($\beta=-.17, p=.043$)の3つが有意であった(表7を参照)。一方、因果強度の説明率は有意に達しなかった($\beta=.06, p=.441$)。因果強度の効果がみられなかったことについては後述し、まず有意であった交互作用項の解釈を行う。

■ 表5：GOAスコアの記述統計

	リンク検索			
	妥当 ($n=68$)		非妥当 ($n=52$)	
構造距離	M	SD	M	SD
近 ($n=67$)	3.57	0.75	1.79	0.69
遠 ($n=53$)	2.38	0.60	1.42	0.28
合計 ($N=120$)	3.06	0.91	1.62	0.57

(注) 因果強度は0から1の値を取る連続尺度として扱ったため、この表からは省略した。

具体的に述べるとリンク×距離の交互作用は、妥当解と非妥当解で構造距離による影響が異なっていたことを示唆している。この点について下位検定を行ったところ、妥当解においては、質問からの距離に伴ってGOAスコアは有意に低下していたが($\beta=-.48, p<.001$)、非妥当解においては、そのような構造距離による影響は有意ではなかった($\beta=-.15, p=.255$)。すなわち、構造距離はリンク検索をパスした妥当解に対してより強く作用しており、よってこのメカニズムはリンク検索によって解答が収束された後に機能したと推測される。ゆえに、この結果から、リンク検索→構造距離という順序を仮定できる。

以上のリンク検索と構造距離による影響の一方、因果強度は有意な貢献を示さず、質問との因果的な

■ 表 6 : GOA スコアと収束メカニズム, およびその交互作用との相関関係

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. GOA	2.44	1.06	-							
2. L	.57	.50	.68**	-						
3. D	.44	.50	-.42**	-.04	-					
4. C	.49	.31	.36**	.12	-.63**	-				
5. L×D	.24	.43	-.03	.49**	.64**	-.37**	-			
6. L×C	.02	.24	.37**	.07	-.48**	.78**	-.53**	-		
7. D×C	-.10	.16	.34**	.06	-.67**	.71**	-.38**	.56**	-	
8. L×D×C	-.05	.13	.06	-.33*	-.42**	.49**	-.66**	.66**	.71**	-

(注) L = リンク検索, D = 構造距離, C = 因果強度, ** $p < .01$

■ 表 7 : GOA スコアに対する階層的回帰分析の結果

ステップと独立変数	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2
ステップ1:					.62**	
L	1.40	0.12	.66**	.000		
D	-0.75	0.16	-.36**	.000		
C	0.20	0.26	.06	.441		
ステップ2:					.66**	.04**
L×D	-0.71	0.35	-.29*	.043		
L×C	0.79	0.57	.16	.227		
D×C	0.74	0.97	.11	.446		
L×D×C	-0.57	1.20	-.07	.637		

(注) L = リンク検索, D = 構造距離, C = 因果強度, * $p < .05$, ** $p < .01$

結びつきの強さは、解答の適切さにあまり影響を与えていなかったことが示唆された。先行研究において、短い説明文を用いた Graesser and Hemphill (1991) でも同様の結果が報告されており、これらを合わせると、テキスト情報間の因果的関連性を正確に反映させた理解を構築することは、読み手に先行知識が比較的少ない説明文に対しては難しかった可能性が考えられる。具体的には、因果的一貫性の高い説明文理解を構築することは、母語読解においても読み手の先行知識量に大きく依存することが知られている (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996)。本調査が対象とした協力者の中に宇宙科学を専門とする者はおらず、イベント間の因果関係の強さを理解に反映させることは、必ずしも容易ではなかったと考えられる。このように、因果強度の GOA に対する影響はあまりみられなかったことから、以降の検証では因果強度を主な考慮から外すこととした。

4.4 調査 1 のまとめ

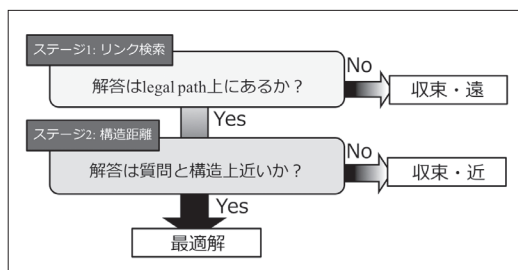
調査 1 では、説明文を読んだ日本語母語話者の質問応答において、3つの解答収束メカニズム (i.e., リンク検索, 構造距離, 因果強度) が働く順番を検証した。GOA スコアに対する重回帰分析から得られた最も重要な結果は、リンク検索と構造距離による交互作用である。すなわち、妥当解の GOA は質問からの距離に従って減少していたが、非妥当解の GOA は距離によってあまり変化しなかった。このことに基づくと、本調査の質問応答に対して表 8 に示されるような解答収束ステージを仮定できる。

■ 表 8 : 調査 1 から示された解答収束ステージ

ステージ 1	リンク検索: 非妥当解から妥当解を区別
ステージ 2	構造距離: 妥当解のうち、質問ノードに近いものへ収束→最適解

調査 2 ではこれを利用して、収束メカニズム上において最適解 (= 正答) に近い解答と、遠い解答を

作成する（図2を参照）。そして、それらの錯乱肢としての有用性を明らかにすることをめざす。



▶ 図2：調査1から示された解答収束ステージと対応する解答

5

調査2

5.1 目的

調査2の目的は、QUESTの収束メカニズムに基づいて作成された錯乱肢が、テキスト理解度の異なる受験者の弁別にどの程度有効かを明らかにすることである。そのために、調査1で用いたテキストに対し、QUESTを利用した錯乱肢2つを含む4択の多肢選択問題を作成した。協力者のテキスト理解は、読んだ内容を覚えている限り書き出す「筆記再生課題」と、テキスト内容に対する原因と結果を因果的に正しい順番に並べて説明する「因果質問課題」により測定した。この2つを用いた理由は、異なるレベルのテキスト理解を対象にするためである。すなわち、読んだことをそのまま書き出す筆記再生課題では、テキストに明示される情報の記憶に依拠する理解が対象となる（以下、明示的理解と述べる；Kintsch, 1994）。一方、因果質問課題では、テキスト内容を一貫した形で結びつけ、因果的に正しい順番で説明することが求められる（5.2.2節を参照；McCrudden et al., 2009）。そのため、テキスト内容に対する構造的な理解が対象となる。これらの課題の成績に基づいて協力者をテキスト理解度上下群にグルーピングし、各選択肢の選択率がテキスト理解度によって異なるか否かを検証する。

5.2 方法

5.2.1 協力者

調査1と同様の国立大学に通う43名の日本人大学生・大学院生が参加した。協力者の専攻は多岐

にわたる。

5.2.2 マテリアル

(1) テキスト

調査1と同じ説明文を英語で用いた。表9に実際のテキストを示す。258語、15文から成り、読みやすさを示すFlesh-Kincaid Grade levelは8.9であった。この指標はアメリカの学年に基づいており、アメリカ人の中学3年生にふさわしい読みやすさと解釈される。また、筆記再生課題の採点に使用する目的で、Ikeno (1996)の基準に従い、調査者と英語教育学を専攻する大学院生の2名によってテキストを78のアイディアユニット（IU）に分割した。分割の一致率は96%で、不一致点は協議により解決した。

(2) 多肢選択問題

テキスト中のイベントをWhyで質問する多肢選択問題を5問作成した。問題と選択肢の例を表9に示す。この例のとおり、各問題にはBecauseで答える選択肢4つが伴った。まず正答の選択肢は調査1で示された2つの収束ステージ（図2を参照）の両方をパスし、QUESTから理論上最適解と想定されるものであった。すなわち、CGSにおいて、質問ノードからのlegal path上にあり（i.e., リンク検索をパス：ステージ1）、かつ質問ノードと隣接している（i.e., 構造距離をパス：ステージ2）解答が該当した。例として、表9のWhy do space travelers become less thirsty?という質問を考えてみたい。これはSN13へのWhy-Event質問で、legal pathはCO, O, R, IN, IMのリンクを逆行するものである（表2を参照）。このlegal pathの上にはSN9を除く、SN3～12のノードがある。このうち、最適解に該当するのは、SN13と隣接するSN12（Some organs try to reduce the amount of blood and water in the upper body.）である。

次に、錯乱肢として、ステージ1のリンク検索はパスするが、質問からの構造距離が遠いためステージ2で収束から外れると想定される解答を、収束メカニズム上最適解に近い解答として用意した。以下、この解答を「収束・近」と述べる。より具体的に、収束・近は最適解と同じlegal path上にあるが、3つ以上のリンクにより質問ノードから離れているSNから作成した。上記Why do space travelers become less thirsty?の収束・近に当たるのは、

■ 表 9：英文と多肢選択式問題の例

When people first considered space travel, they did not know how the zero gravity of space would affect humans. The human body is a complex system that automatically responds to the lack of gravity.

While in space, the body is not affected by gravity. Therefore, blood and water do not travel to the lower parts of the body, especially the legs. Instead, the blood and water within the body move to the upper body. This leads to some interesting effects. Because the blood and water travel to the upper parts of the body, the body feels like the chest and head are filled with blood and water. Because of this, the heart and lungs send messages that the amount of blood and water in the upper part of the body must be reduced. As a result, space travelers do not feel thirsty, and therefore drink less water. As body water is eliminated, their body water levels become lower than normal. When the amounts of blood and water decrease, it becomes more difficult for the human body to work normally. Therefore, the heart does not work as hard as it does on Earth. As a result, the heart becomes smaller.

Successful human research in space depends on understanding how the human body is influenced by the environment in outer space. Another benefit of this research is that studying the effects of space travel on the human body can help us better understand many illnesses suffered by people living on Earth, such as high blood pressure and other heart problems.

Why do space travelers become less thirsty?

- (a) Because organs try to lower the amount of water in the body. (最適解)
- (b) Because the human body is not influenced by gravity. (収束・近)
- (c) Because space travelers drink less water than normal. (収束・遠)
- (d) Because the lack of gravity makes the nervous system less sensitive. (Non-Text)

- (a) 最適解：質問ノードからの legal path 上にあり、質問ノードと構造的に隣接する解答
- (b) 収束・近：legal path 上にあるが、質問ノードから構造的に遠い解答
- (c) 収束・遠：legal path と反対方向のリンクによって質問ノードと構造的に隣接する解答
- (d) Non-Text：テキストに述べられていない情報による解答

(注) 調査では選択肢はランダム提示された。

legal path 上にある SN (i.e., SN3, 4, 5, 6, 7, 8, 10, 11, 12) のうち、SN13からリンク5つ分離れている SN5 (Gravity does not affect the human in space.) である。

そして、ステージ1のリンク検索の時点で収束から外れる解答を、収束メカニズム上最適解から遠い解答として作成した。この解答は「収束・遠」と述べる。収束・遠には、legal path と反対方向のリンクによって質問ノードと隣接する SN を用いた (cf. Millis & Barker, 1996)。上記 SN13の場合、legal path の反対方向のリンク (順行の Initiate) により隣接する SN14 (Space travelers drink less water.) が該当する。

最後に、比較対象として、テキストにない情報による錯乱肢を作成した。これは便宜上、Non-Text と述べる。Haladyna and Downing (1989) の錯乱肢作成ガイドラインに従い、テキストで使用されている語句を用い、科学的な事実に基づくよう留意した。

これらの選択肢の作成においては、テキスト内容の字義的な記憶による影響が大きくなりすぎないよ

うに、可能な限り原文からのパラフレーズを使用した。質問と選択肢は、ネイティブスピーカーによって表現や意味に不自然がないかを確認され、必要な場合、調査者とネイティブスピーカーの2名により微修正が加えられた。

(3) 筆記再生課題

筆記再生課題は紙面にて作成され、以下の3つの注意点が与えられた：(a) 英文の内容で覚えていることを残らず書き出すこと、(b) 単語を羅列するのではなく、できるだけ文章の形で書くこと、(c) 繰り返しや余分なことがあっても構わないのでできるだけ多くのことを書くこと。またテキストは1つのみで、かつ読解直後に行われたことから、課題時に再生の手がかりは与えられなかった。

(4) 因果質問課題

因果質問課題も紙面にて作成された。この課題では、テキスト中のイベントに対する原因や結果を、適切な順番に並べて説明することが求められる

(McCrudden et al., 2009; Ushiro et al., 2015)。本調査のテキストは「無重力空間にいること」が「心臓が小さくなること」を引き起こすまでの過程を因果関係に基づき記述しているが、因果質問課題はそれを大きく前半と後半に分けて説明することを求めた(資料3)。

まず前半では「無重力空間にいること」が、「体が水分を減らそうとする」のはなぜかを、因果的に正しい順番で、可能な限り詳しく説明することが求められた。続く後半は、「体が水分を減らそう」とした結果、人体に何が起きるとテキストに書いてあったかを、因果的に正しい順番で詳しく説明することが求められた。また、協力者が課題の方法や形式に慣れられるよう、テキスト内容とは異なるイベントを用いた例が与えられた。

5.2.3 手順

調査は個別に実施され、およそ1時間で終了した。協力者はテキストを読解した後、筆記再生課題、因果質問課題、多肢選択問題の順に取り組んだ。多肢選択問題の際には、質問に聞かれていることに対し、テキストの内容に基づいて、最も正しく、最も適切で、かつ最も直接的な答えを選ぶよう指示が与えられた。また、順序による影響をなくすために、選択肢はランダムに並び順で提示された。

なお注意すべき点として、解答収束メカニズムは記憶内の知識構造に基づくことから (Graesser & Clark, 1985)、筆記再生以降の課題はすべて、テキストの参照を許可しない記憶ベースの形式をとった (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007)。この影響については後述する。

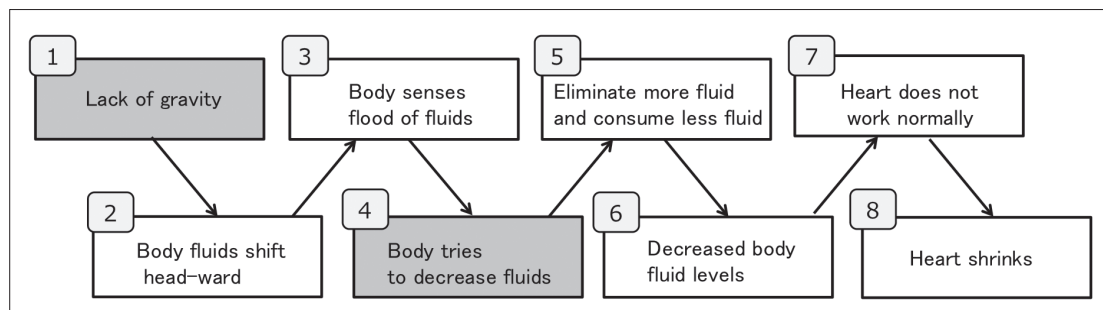
5.2.4 採点・分析

多肢選択問題について、各選択肢が選ばれた割合

を算出した。筆記再生課題に関しては、各協力者のIU再生率を計算した^(注6)。採点では、各IUのおよそ3分の2に相当する情報が再生されていれば、そのIUに1点を与えたが、科学的に誤った記述や、テキストになかった記述には点を与えなかった。最初に、全データの30%を調査者と英語教育学を専攻する大学院生の2名が採点し(一致率86%)、不一致点を協議より解決した。その上で採点基準を明確化し、残りのデータは調査者が1名で採点した。

因果質問課題に関しては、Ushiro et al. (2015)がMcCrudden et al. (2009)に基づいて作成した「因果ダイアグラム」を利用して採点した。図3のとおり、因果ダイアグラムには、宇宙空間の「無重力」が「心臓が小さくなる」ことを引き起こすまでに経る8つの命題 (Proposition: 以下、Pと述べる) が示されている。

このうち、解答時に与えた2つ (i.e., P1, P4) を除く6命題に該当する記述で、解答中において因果的に正しい順番で書かれているものに1点ずつを与えた (Min/Max=0/6)。採点においては、因果的な順序が正しければ、途中のイベントが抜けていても点を与えたが、順序が逆転していたり、科学的に誤った記述には点を与えなかった。例として、解答の後半部分に相当する「(体が水分を減らそうとした結果) 喉が渇かなくなって (P5)、水を飲まなくなる (P6)。そして心臓が小さくなって (P8)、あまり動かなくなる (P7)」という記述を考えてみたい。ここでは、P7が本来後に起こるP8の前に書かれているため、それを除く命題を合計した3点が与えられる (P5+P6+P8)。この基準に従い、全データの30%を調査者と英語教育学を専攻する大学院生の2名が採点した(一致率90%)。不一致点は協議によって解決し、残りのデータは調査者が1名で採点した。



▶ 図3：因果ダイアグラム

5.3 結果と考察

まず、協力者全体における選択肢の選択率の傾向をみて、その後、テキスト理解度と各選択肢の選択率のかかわりを調べる。

表10に、各選択肢が選ばれた割合を示す。これら4つの選択率を一元配置分散分析 (analysis of variance: ANOVA) によって比較したところ、最適解の選択率が最も高く、Non-Text が最も低かった (all p s < .05)。一方、収束・近と収束・遠の選択率に有意な差は確認されなかった (p = .195)。この結果から、全体的な傾向として、(a) QUEST の最適解は、正答と判断される確率が十分に高いこと、(b) 収束メカニズムに基づく解答 (i.e., 収束・近、遠) は、テキストにない情報 (i.e., Non-Text) よりも受験者を多く引きつけたことがわかった。続いて、これらの選択率が、テキスト理解の度合いによってどのように変化したかを検討する。

■ 表10：各選択肢が選ばれた割合 ($N = 43$)

最適解		収束・近		収束・遠		Non-Text	
M	SD	M	SD	M	SD	M	SD
.70	.23	.16	.19	.11	.15	.04	.08

筆記再生、因果質問課題の成績をそれぞれ表11、12に示す。まず、テキスト理解の異なる協力者を区別する目的で、課題の成績に基づくグルーピングを行った。ここで、本調査の課題の成績では度数や標準偏差が比較的小さかったため、中央値によりグループを分割すると、カットオフスコア付近ではほとんど同じ成績の協力者が異なるグループに分かれてしまう問題があった (Pelham & Blanton, 2012)。そこで、対象となる人数は少なくなってしまうが、筆記再生課題では、再生率上位33%を上位群 ($n = 12$)、下位33%を下位群 ($n = 12$)、その間を中位群 ($n = 19$) とし、うち上下群を分析対象とした。因果質問課題においては、5、6点を得た者を上位群 (n

■ 表11：筆記再生課題の結果

	n	M	SD	Min	Max
上位	12	.45	.08	.41	.66
中位	19	.34	.04	.28	.38
下位	12	.14	.06	.03	.22
合計	43	.31	.13	.03	.66

(注) 上位群と下位群を分析の対象とした。

■ 表12：因果質問課題の結果

	n	M	SD	Min	Max
上位	14	5.36	0.50	5.00	6.00
中位	18	3.56	0.51	3.00	4.00
下位	11	1.18	0.87	0.00	2.00
合計	43	3.53	1.71	0.00	6.00

(注) 上位群と下位群を分析の対象とした。

=14)、3、4点を中位群 ($n=18$)、0～2点を下位群 ($n=11$) とし、同じく上下群を分析対象とした。

この上下群間において、選択肢の選ばれ方がどのように異なっていたかを調べるために、選択肢の選択率を従属変数とした2 (テキスト理解：上位、下位) × 4 (選択肢：最適解、収束・近、収束・遠、Non-Text) の2要因混合計画のANOVAを行った。グループが協力者間要因、選択肢が協力者内要因であった。またこのANOVAは、筆記再生課題に基づく明示的理解によるグループと、因果質問課題に基づく構造的 understanding によるグループに対し個別に行った。

明示的理解の上下群における各選択肢の選択率を表13に、構造的 understanding の上下群における選択率を表14に示す。分析の結果、両テキスト理解に対するANOVAにおいて、選択肢の主効果、明示： $F(3, 66) = 68.01, p < .001, \eta^2 = .76$ ；構造： $F(3, 69) = 57.73, p < .001, \eta^2 = .72$ 、および、テキスト理解 × 選択肢の交互作用、明示： $F(3, 66) = 68.01, p < .001, \eta^2 = .76$ ；構造： $F(3, 69) = 57.73, p < .001, \eta^2 = .72$

■ 表13：明示的理解 (筆記再生課題に基づく) の上下群における各選択肢の選択率

	最適解		収束・近		収束・遠		Non-Text	
	M	SD	M	SD	M	SD	M	SD
上位 ($n=12$)	.85	.15	.08	.13	.05	.09	.02	.06
下位 ($n=12$)	.55	.26	.28	.23	.12	.13	.05	.09
合計 ($n=24$)	.70	.26	.18	.21	.08	.12	.03	.08

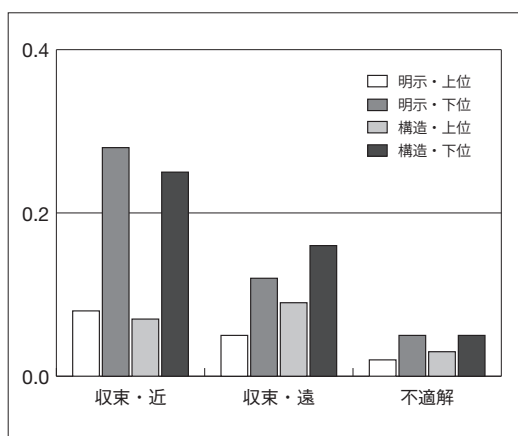
■ 表14：構造的 understanding (因果質問課題に基づく) の上下群における各選択肢の選択率

	最適解		収束・近		収束・遠		Non-Text	
	M	SD	M	SD	M	SD	M	SD
上位 ($n=14$)	.81	.18	.07	.10	.09	.10	.03	.07
下位 ($n=11$)	.53	.26	.25	.22	.16	.20	.05	.09
合計 ($n=25$)	.69	.26	.15	.19	.12	.15	.04	.08

が有意であった。

ここからは、有意であったテキスト理解×選択肢の交互作用の解釈を行う。具体的に、この交互作用は、(a) テキスト理解度による選択率の違いは、特定の選択肢に限られていたこと、(b) 選択肢間における選択率の違いは、テキスト理解度によって異なっていたことを意味している。

まず、(a)について下位検定を行った結果、両レベルのテキスト理解 (i.e., 明示, 構造) において、上位群は下位群より有意に多く最適解を正答に選んでいたことがわかった (both $ps < .05$)。一方、錯乱肢 (i.e., 収束・近, 収束・遠, Non-Text) については、両レベルのテキスト理解において、下位群は上位群より有意に多く収束・近を選択していた (both $ps < .05$)。一方、収束・遠と Non-Text に関しては、テキスト理解度による選択率の差は確認されなかった (all $ps > .100$)。よって、この結果から、収束メカニズム上最適解に近い錯乱肢 (i.e., 収束・近) は、テキストの理解度が低い学習者をより多く引きつけたことがわかった (図4を参照)。



▶ 図4：明示的・構造的理解の上下群における錯乱肢の選択率

次に、(b)について、上位群では最適解が他のどの選択肢よりも有意に多く選ばれていたが (all $ps < .05$)、下位群では、最適解と収束・近の選択率に統計的有意差は確認されなかった (both $ps > .100$)。この結果にテキスト理解のレベル (i.e., 明示, 構造) による違いはなく、収束・近は下位群を多く引きつけていたことが支持された。

さらに、ANOVAでは対象とするデータを限定したことから、結果を補完する目的で、全データを対

象にテキスト理解度 (i.e., 筆記再生課題, 因果質問課題の成績) と各選択肢の選択率との相関分析を行った。その結果、表15に示されるとおり、最適解の選択率と両テキスト理解とは中程度の正の相関があり (明示： $r=.55$, 構造： $r=.51$ (注7))、最適解を多く正答に選んだ協力者は、テキスト理解が優れていた傾向が確認された。一方、収束・近については、弱い～中程度の負の相関が、明示的 ($r=-.47$)、構造的な理解 ($r=-.36$) に対しそれぞれ得られ、収束・近を多く選んだ協力者は、テキスト理解が低い傾向にあったことが示唆された。ここで明示的理解の方が構造的な理解よりもやや高い相関係数を示しているが、統計的有意差は確認されなかった ($t=1.02$, $p=.312$)。

■ 表15：各テキスト理解度と各選択肢の選択率の相関

	明示的理解 (筆記再生課題に 基づく)	構造的な理解 (因果質問課題に 基づく)
最適解	.55**	.51**
収束・近	-.47**	-.36*
収束・遠	-.14	-.23
Non-Text	-.17	-.20

(注) $N = 43$, * $p < .05$, ** $p < .01$

このように、テキスト理解度は最適解の選択率と正の相関、収束・近の選択率とは負の相関関係にあることがわかった。ANOVAの結果とあわせると、テキスト理解度が低い受験者は、妥当解 (i.e., 最適解, 収束・近) と非妥当解 (i.e., 収束・遠) とは区別できるものの、構造距離による解答収束でつまづいたことが示唆される。これに対する理由として、そのような受験者は、心内でのテキスト情報のネットワーク構造があまり精緻でなく、ゆえに図2のステージ2において最適解へ収束させるとき、構造上遠い解答と混乱が生じたためであることが考えられる。ただし、結果について注意に値する点が2つある。1点目は、テキスト理解のレベル (i.e., 明示的, 構造的な理解) による違いがみられなかったことである。具体的には、テキスト情報によるネットワークの精緻さは、テキストに明示される情報の再生量で決まる明示的理解 (筆記再生課題により測定) よりも、情報間の因果的なつながりが問われる構造的な理解 (因果質問課題により測定) とより強く関係していると考えられる (e.g., Kintsch, 1994; McCrudden

et al., 2009; McNamara et al., 1996)。これについては、本調査で用いたテキストの性質による可能性が考えられる。すなわち、実験テキストは、無重力が人体に与える影響を、原因と結果の順序に従って簡潔に述べたものであった。この事象の因果的な展開に関する記述以外はあまりなく、よって、テキストに明示的に述べられている内容を理解すれば、その因果的構造がある程度把握できる性質であったと考えられる。実際に、筆記再生と因果質問課題の成績は中程度の相関にあり ($r=.68$)、2つの理解のグルーピングにバリエーションはあまりなかった。このことから、本調査のテキストにおいては、明示的理解と構造的理解とが比較的近接したものであったと考えられる。

2点目は、多肢選択問題を記憶ベース (i.e., 課題遂行時にテキストの参照を許可しない) の形式で行った点である。Ozuru, Dempsey, and McNamara (2009) によると、記憶ベース形式は、テストに向けて教科書を学習するが、教科書の参照はテスト時には普通許可されないという生徒の環境を反映する利点がある。一方で、英検やTOEICなどの既存の読解テストでは、テキストと選択肢が同時に与えられるのが普通である。この場合、受験者は選択肢から得られる手がかりを基に、テキスト中の関連情報を検索するような付加的なストラテジーをとることが知られている (Ozuru et al., 2007)。ただし、QUEST が想定する範囲はテキストから理解された内容による知識構造であって、このような方略的処理を既存の解答収束メカニズムによって予測することは難しい。受験者の方略的処理を考慮したモデルの開発は、検証価値のある課題と言えよう。

5.4 調査2のまとめ

調査2では、QUESTの解答収束メカニズムに基づいて作成された錯乱肢が、テキスト理解度の異なる学習者をどの程度弁別するかを検証した。その結果、収束メカニズム上において最適解に近い錯乱肢 (i.e., 収束・近) は、テキスト理解度が低い学習者をより多く引きつけることがわかった。一方、テキスト中であっても最適解から遠い錯乱肢や、テキストにない情報の錯乱肢は、選択率にテキスト理解度による違いはみられなかった。この結果から、もっともらしい錯乱肢の条件として、答えとして完全に間違いではないが (i.e., 妥当解)、最適かどうかを

問われたとき、十分に直接的ではない (i.e., 構造距離が遠い) ということが挙げられる。少なくとも本研究で用いたような Why 型の設問については、これを錯乱肢作成の1つの目安になると考えられる。

6

結論と今後の課題

本研究は、2つの調査を通して、質問応答モデル QUEST を多肢選択式読解テストにおける錯乱肢作成に利用する可能性を検証した。以下に主要な結果をまとめる。

調査1では、QUEST が想定する3つの解答収束メカニズム (i.e., リンク検索、構造距離、因果強度) がどのような順番で働くのかを調べた。その結果、まずリンク検索によって妥当解が特定され、次に妥当解の中で質問に構造距離が近いものへ収束するという順序が確認された。そして続く調査2では、調査1から示唆された解答収束ステージを利用して、最適解に近い収束メカニズムに近い、および遠い錯乱肢を作成し、それらがテキスト理解度の異なる受験者の弁別にどの程度有効かを調べた。結果、テキスト理解度の低い受験者は、テキスト理解度が高い受験者と比べ、最適解に近い錯乱肢 (i.e., 構造距離が遠い妥当解) を多く選択していた。ゆえに、この錯乱肢が、テキスト理解度の異なる学習者の弁別に有用である可能性が示された。

これらの調査を通し、本研究は英文読解テストに対する質問応答モデルの応用可能性を示したものである。従来、質問応答にかかわる研究成果は、質問紙やコンピュータチューターの作成に活用されてきたものの、それを外国語教育へ応用する取り組みは決して十分ではなかった。もっともらしい錯乱肢の作成に対する可能性のあるアプローチは、高木 (2011) が指摘するよう、学習者の記憶や認知面を観点とすることである。本研究は、QUEST の CGS と解答収束メカニズムによってこの点に取り組むことを通し、質問応答モデルと言語テストとをつなぐ1つのきっかけを作ったといえる。

その一方で、本研究の結果にはいくつかの留意点を指摘する必要がある。まず大きな点として、使用した項目やテキスト、および対象とした協力者の数が非常に少なかったことが挙げられる。このため、結果をどの程度一般化できるかには注意が必要であ

る。例えばテキストに関して、大学などの高等教育機関における英語教育では、より複雑で抽象的な内容の英文をテスト材料として用いる場合がある。そのようなテキストでは、最初の収束ステージであるリンク検索すら困難になるかもしれない。もしそうならば、リンク検索に違反する非妥当解 (i.e., 調査2における収束・遠; e.g., Why 質問に対し結果で解答する) が、有効な錯乱肢として機能する可能性があるだろう。さらに、記憶ベースの解答形式 (i.e., 解答時にテキストを参照しない) をとった点も留意に値する。読解テストでは、テキスト、設問、選択肢がセットで与えられる場合が多く、このとき、受験者は選択肢を手がかりとして関連情報をテキスト中に検索するなどのストラテジーをとることが知られている (e.g., Ozuru et al., 2007)。残念なことに、QUEST が想定するのはテキスト内容を基に構築された知識構造のみであるため、本研究で用いた解答収束メカニズムによってこのような付加的な方略的処理を説明することは難しい。テスト環境での質問応答に最適化されたモデルの開発は、今後大いに検証が期待される課題である。最後に、実用面に関する留意点を指摘できる。QUEST を利用するにはテキストの理解表象を反映した CGS が必要だが、現状、これは調査1に記されるような煩雑な作業を強いられる。本研究の範囲を超えるものだが、与えられたテキストに対して、自動的に CGS や各質問へ

の妥当解を出力するようなプログラムの開発が望まれる。

これらの点を考慮すると、本研究は質問応答モデルを言語テストに利用した、あくまで最初のステップとしてとらえるのが妥当であろう。今後は、この分野の研究知見から示唆される、人が質問に対してとる認知的処理や方略を、テスト作成の文脈へと具体化していく試みが望まれる。それによって、テスト作成に盤石な基盤が形成され、より正確な能力推定が可能になるであろう。

謝 辞

本研究を実施する貴重な機会を与えてくださいました公益財団法人 日本英語検定協会の皆さま、ならびに選考委員の先生方に厚く御礼申し上げます。特に、私の研究を担当してくださいました大友賢二先生には、たくさんの有益なご助言・ご指導をいただき、心より御礼申し上げます。そして、筑波大学大学院の卯城祐司先生には、調査の実施から報告書執筆に至るまで親身なご指導をいただきました。心より感謝申し上げます。また、研究室の同期である森好紳さんには、実験材料の作成を中心にお手伝いいただき、いつも的確なアドバイスをいただきました。最後に、調査にご協力いただきました、協力者の皆さまに深く御礼申し上げます。

注

- (1) 厳密に述べると、高木 (2011) で考慮された因果的連続性と、QUEST が想定する因果強度は異なる。因果的連続性は、原因 A と結果 B 間における十分性により因果関係を規定するが (i.e., A が起これば B も起こる)、QUEST の因果強度は、十分性に加え必要性 (i.e., A が起きなければ B も起きない)、時間性 (A は B よりも前に起こる)、操作性 (B が起これば A も起これば) が考慮される。
- (2) 本稿では簡潔さをとって Why 質問の結果のみを記したが、他の質問タイプ (e.g., how, when) については、このような順序性が確認されなかったものもあった。
- (3) カッパ係数は、順序尺度または順序尺度のデータに適用される信頼性の算出手法である。値が .00 ~ .20 のとき信頼性は「低い」、.21 ~ .40 は「やや低い」、.41 ~ .60 は「中程度」、.61 ~ .80 は「かなり高い」、.81 ~ 1.00 は「ほぼ一致」と解釈される (平井, 2012, p.151)
- (4) 行動を why で聞いた場合、物理的な原因だけでなくその行動の目的も妥当な解答となる (Graesser & Hemphill, 1991)。例えば、「なぜ部屋を掃除するの

- か」という質問に対しては、「部屋が散らかっているから (原因)」に加え「物を整理するために (目的)」も妥当な解答となる。
- (5) 因果強度の算出方法を具体的に述べると、情報 AB 間の時間性 (temporality [T]: A が B より時系列的に前に起これば)、操作性 (operativity [O]: B が起これば A も起これば)、必要性 (necessity [N]: A が起これば B も起これば)、十分性 (sufficiency [S]: A が起これば自然に B も起これば) を 0, 0.5, 1 の 3 段階で評価し、次の式に代入する (van den Broek, 1990), $\text{Temporality} \times \text{Operativity} \times (\text{Necessity} + \text{Sufficiency}) / 2$
- (6) 親近性効果を避けるために、最終段落に相当する IU 67 ~ 78 は分析対象外とした (Ushiro et al., 2015)。
- (7) 相関係数の値は $-1 \leq r \leq 1$ の範囲をとり、.00 ~ ±.20 で「ほとんど相関がない」、±.20 ~ ±.40 で「弱い相関がある」、±.40 ~ ±.70 で「中程度の相関がある」、±.70 ~ ±1.00 で「強い相関がある」と解釈される (平井, 2012, p.139)。

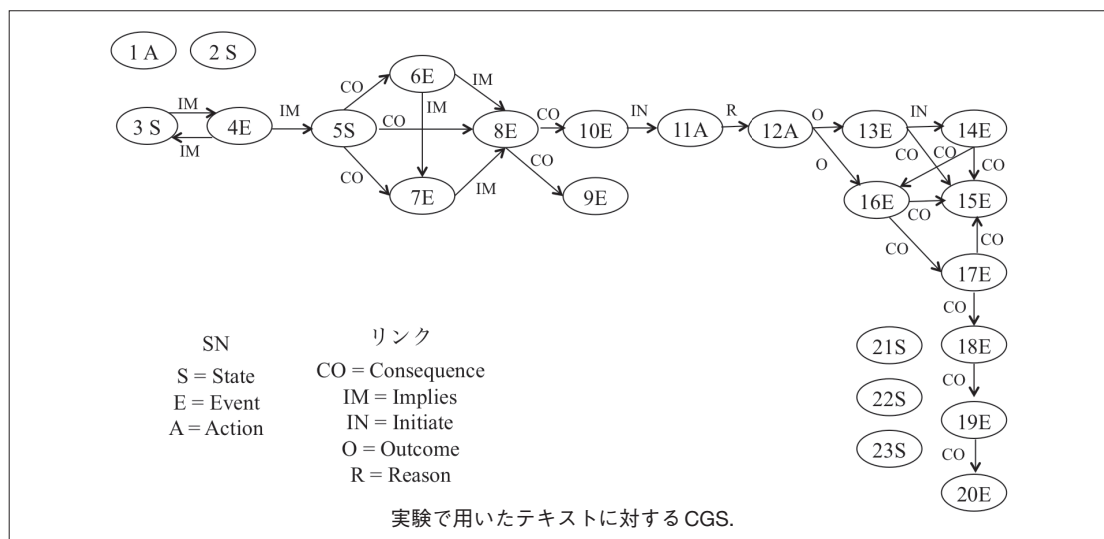
参考文献 (*は引用文献)

- * 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.
- * Graesser, A.C., & Clark, L.F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- * Graesser, A.C., & Franklin, S.P. (1990). QUEST: A cognitive model of question answering. *Discourse processes*, 13, 279-303.
- * Graesser, A.C., & Hemphill, D. (1991). Question answering in the context of scientific mechanisms. *Journal of Memory and Language*, 30, 186-209.
- * Graesser, A.C., Lang, K.L., & Roberts, R.M. (1991). Question answering in the context of stories. *Journal of Experimental psychology: General*, 120, 254-277.
- * Graesser, A.C., Roberts, R.M., & Hackett-Renner, C. (1990). Question answering in the context of telephone surveys, business interactions, and interviews. *Discourse Processes*, 13, 327-348.
- * Graesser, A.C., VanLehn, K., Rosé, C.P., Jordan, P.W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI magazine*, 22, 39-52.
- * 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., 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.
- * 平井明代 (編著). (2012). 『教育・心理系研究のためのデータ分析入門：理論と実践から学ぶ SPSS 活用方法』. 東京：東京図書.
- * Ikeno, O. (1996). The effects of text-structure-guiding questions on comprehension of texts with varying linguistic difficulties. *JACET Bulletin*, 27, 51-68.
- * Kintsch, W. (1994). Text comprehension, memory, and learning. *American psychologist*, 49, 294-303.
- * McCrudden, M.T., Schraw, G., & Lehman, S. (2009). The use of adjunct displays to facilitate comprehension of causal relationships in expository text. *Instructional Science*, 37, 65-86.
- * McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, 14, 1-43.
- * Millis, K.K., & Barker, G.P. (1996). Answering questions about expository texts. *Discourse Processes*, 21, 57-84.
- * Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25, 399-438.
- * Ozuru, Y., Dempsey, K., & McNamara, D.S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and instruction*, 19, 228-242.
- * Pelham, B.W., & Blanton, H. (2012). *Conducting research in psychology: measuring the weight of smoke*. Belmont: Wadsworth.
- * 高木修一. (2011). 「多肢選択式問題は公平か—記憶保持の強さから—」. *EIKEN BULLETIN*, vol.23, 13-29.
- * Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, 16, 3-34.
- * Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24, 612-630.
- * Ushiro, Y., Hamada, A., Kimura, Y., Nahatame, S., Hosoda, M., Kato, D., & Watanabe, Y. (2015). Building causally coherent mental representations and learning from expository texts among Japanese EFL readers. *JACET Journal*, 59, 131-150.
- * 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., Kiyama, M., Suzuki, R., Kurisu, Y., & Gomi, N. (2007). What makes distractors plausible in multiple-choice reading tests? *JLTA Journal*, 10, 56-67.
- * van den Broek, P. (1990). Causal inferences and the comprehension of narrative text. In A.C. Graesser & G.H. Bower (Eds.), *The psychology of Learning and Motivation: Inferences and text comprehension*. New York: Academic Press.
- * Zwaan, R.A., Langston, M.C., & Graesser, A.C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *psychological science*, 6, 292-297.

資料 1：本調査で用いたテキストに対する SN と CGS

SNの一覧

番号		種類
SN1	People considered space travel.	Action
SN2	People did not know effects of zero gravity on humans.	State
SN3	The human body is a complex system.	State
SN4	The human body automatically responds to the lack of gravity.	Event
SN5	Gravity does not affect the human in space.	State
SN6	The blood and water do not travel to the lower body.	Event
SN7	The blood and water do not travel to legs.	Event
SN8	The blood and water move to the upper body.	Event
SN9	Some interesting effects occur.	Event
SN10	The body feels like the chest and head are filled with blood and water.	Event
SN11	The lungs and heart send messages.	Action
SN12	Some organs try to reduce the amount of blood and water in the upper body.	Action
SN13	Space travelers do not feel thirsty.	Event
SN14	Space travelers drink less water.	Action
SN15	The body water level becomes less than normal.	Event
SN16	The body water is eliminated.	Event
SN17	The amount of blood and water decreases.	Event
SN18	The human body has difficulty in working normally.	Event
SN19	The heart does not have to work as hard as on Earth.	Event
SN20	The heart becomes smaller.	Event
SN21	Successful space research depends on the understanding of effects of space on the human body.	State
SN22	Effects of space travel can help us better understand many illnesses.	State
SN23	Many people on Earth suffer from the illnesses.	State



資料 2：調査 2 の多肢選択問題

(注) 調査では、選択肢はランダムに並び順で提示された。

- (1) Why don't blood and water travel to the lower body in space?
 - (a) Because the human body is not influenced by gravity in space. (最適解)
 - (b) Because the human body is a complicated system. (収束・近)
 - (c) Because the human body experiences the increase in water in the upper body. (収束・遠)
 - (d) Because the human body is on average denser than water. (Non-Text)
- (2) Why does the body feel like the chest and head are filled with blood and water in space?
 - (a) Because the body water shifts to the upper body. (最適解)
 - (b) Because the human body makes an automatic response to the lack of gravity. (収束・近)
 - (c) Because organs send messages to parts of the body. (収束・遠)
 - (d) Because the body's muscles become weaker when there is no gravity. (Non-Text)
- (3) Why do some organs try to reduce the amount of water in the upper body in space?
 - (a) Because the human body experiences an increase in water in the upper body. (最適解)
 - (b) Because the human body is a complicated system. (収束・近)
 - (c) Because there is less need for the heart to work hard in space. (収束・遠)
 - (d) Because the lower part of the human body becomes smaller in space. (Non-Text)
- (4) Why do space travelers become less thirsty?
 - (a) Because the organs try to lower the amount of water in the body. (最適解)
 - (b) Because the human body is not influenced by gravity. (収束・近)
 - (c) Because space travelers drink less water than normal. (収束・遠)
 - (d) Because the lack of gravity makes the nervous system less sensitive. (Non-Text)
- (5) Why does the body's water level become lower than normal in space?
 - (a) Because water is reduced by the body's systems. (最適解)
 - (b) Because the human body experiences an increase in water in the upper body. (収束・近)
 - (c) Because it becomes difficult for the human body to function efficiently. (収束・遠)
 - (d) Because it is less humid in space than on Earth. (Non-Text)

資料 3：調査 2 の因果質問課題

英文の内容を順序立てて説明する課題を行っていただきます。

注意点 原因と結果を、論理的に正しい順番に最初から並べて、可能な限り詳しく説明してください。

【例題】 地震が起こるのはなぜですか。その原因を最初から順序立てて、可能な限り詳しく説明してください。

地球のプレートが移動し、海のプレートと陸のプレートがぶつかり合うと陸のプレ-

ト側が巻き込まれ、巻き込まれたプレートの端が反発することで地震が発生する。

Q 無重力空間にいと、体が水分を減らそうとするのはなぜですか？

また、体が水分を減らそうとした結果、何が起きますか？

原因と結果を出来るだけあげながら、可能な限り詳しく説明してください。

無重力空間にいと、

体が水分を減らそうとする。

その結果、