キーセンテンスの内容と文章中の範囲が 能力別にみた錯乱枝の選べれ方に及ぼす影響

—英語文章読解テストを用いて—

○寺尾 尚大¹, 石井 秀宗¹, 野口 裕之¹

¹ 名古屋大学大学院教育発達科学研究科

問題

テストが受検者の能力を精度よく測定するためには、良問の作成が欠かせない。多枝選択式テストの項目作成ガイドラインによれば、「すべての誤答選択枝がもっともらしくなるように作成すること」(Rule 29)が重要であると指摘されている (Haladyna, Downing & Rodriguez, 2002)。誤答選択枝のうち、もっともらしい誤答選択枝を錯乱枝と呼ぶ (寺尾・安永・石井・野口, 2015)。英語文章読解テストを用いて錯乱枝の効果を検討した研究によれば、能力群ごとに選べれやすい錯乱枝が異なっていた (寺尾ら, 2015)。

受検者は文章を読解する際、設問で問われている内容に対する答えを得るため、設問に対する答えとなる文章中の記述を探す。問題に正答するために参照するヒントとなる文をキーセンテンスと呼ぶ (Carr, 2006)。本研究は、キーセンテンスの種類に着目し、キーセンテンスの種類によって錯乱枝の選べれ方が能力群別に異なるか検討することを目的とする。予備調査ではキーセンテンスの種類に関するカテゴリについて情報収集および整理を行い、本調査では整理したカテゴリをもとに、キーセンテンスの種類の違いによる錯乱枝の選べれ方の違いを能力群別に検討する。

予備調査

目的

キーセンテンスが実際の入学試験問題においてどのように設定されているか検討し、キーセンテンスの種類に関するカテゴリを整理することであった。

方法

研究協力者 心理学を専門とする大学院生 4 名であった。

材料 私立大学で過去に出題された英語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複数用いた。問題の選定に際し、近年出題された入学試験問題であること、使用する問題全体で以下で述べる分類カテゴリを網羅できることに配慮した。

分類カテゴリ キーセンテンスの種類を内容と文章中の範囲の 2 側面から捉え、先行研究を参考にカテゴリを設定した。キーセンテンスの内容については、テストで測定される知識の分類 (Haladyna, 2002) に基づき、「事実」「概念」「手続き」「原理」に加え、筆者の意見が述べられているようなキーセンテンスがあることを考慮に入れ、本研究で独自に「意見」というカテゴリを含めて 5 カテゴリを提示した。

また、キーセンテンスの文章中の範囲については、一杉 (2007) の「部分一致問題」「全体一致問題」をより精緻にし、「文章中の一文を踏まえる」「文章中の一つの段落全体を踏まえる」「文章中の複数の段落を踏まえる」「文章全体を踏まえる」の 4 カテゴリを提示した。カテゴリの定義を Table 1 に示した。

調査冊子の構成 設問幹・正答選択枝に加え、その問題に解答する上で踏まえなければならないキーセンテンスを提示し、これらをまとめて 1 セクションとした。1 冊の調査冊子には、20 セクション前後が収録されるように調整した。提示したキーセンテンスの数は問題に応じて異なっており、一文の場合もあれば、複数の文の場合もあった。研究協力者が当該セクションにおけるキーセンテンスの文章内での散らばり具合を判断できるよう、キーセンテンスの文章中での位置を「第○段落・△文目」と表示した。

手続き 研究協力者としてリクルートした大学院生に、本研究の目的を伝え、研究協力への同意を確認した。次に、調査冊子への回答方法および分類カテゴリについての詳細な説明を行い、2 週間程度で調査冊子に回答するよう求めた。2 週間後、第一著者とそれぞれの研究協力者との 2 名で合議を行い、キーセンテンスの内容および文章中での範囲に関する情報を収集した。また、カテゴリの再定義や名称変更が必要な箇所について提案を求めた。以上のような合議を研究協力者 4 名と行った。

結果

カテゴリの再定義・名称変更 研究協力者との合議の過程で、カテゴリの再定義に関する提案が四点あった。うち二点は、カテゴリの名称変更を含む提

Table 1 調査冊子に提示したキーセンテンスの分類カテゴリ

カテゴリ名	定義	カテゴリ名	定義
キーセンテンスの内容		キーセンテンスの文章中の範囲	
事実	全員が一致できるようなまぎれもない真実について述べたもの	一文を踏まえる	各選択枝のキーセンテンスが一つの文に集中している
概念	定義を必要とし、さまざまな特徴によって区別されるものについて述べたもの	一つの段落を踏まえる	各選択枝のキーセンテンスが同一段落中にある
手続き	人がある目的を達成するときに必ず従う一連の動作、流れについて述べたもの	複数段落を踏まえる	各選択枝のキーセンテンスが複数の段落にまたがっている
原理	あるものやことがらが機能する仕組みについて述べたもの	文章全体を踏まえる	各選択枝のキーセンテンスが文章全体に散らばっている
意見	筆者が文章中に述べていることがらについての意見を述べたもの		

案であった。

第一に、「事実」について「全員が一致できるようなまぎれもない真実について述べたもの」という定義が限定的すぎて、このカテゴリにあてはまるものはなかったと報告する研究協力者がいた。この指摘を踏まえて、カテゴリの定義を「ほとんどの人々が納得できることがらについて述べたもの」に変更した。

第二に、「意見」について、文章中には筆者でない人の意見も記述されていることがあるとの指摘があった。学者の意見や言説が述べられている文がキーセンテンスとなっていた場合には、意見カテゴリに分類することが適切であると判断し、カテゴリの定義を「文章中に述べられている意見」に変更した。

第三に、「手続き」について、時代背景や制度の成立過程などについて述べられているキーセンテンスを、手続きカテゴリに分類する研究協力者がいた。カテゴリの名称を「手続き・過程・背景」とし、定義を「一連の動作・流れ、成立過程や背景について述べたもの」に改めた。

第四に、「一文を踏まえる」について、研究協力者から「4つの選択枝すべてのキーセンテンスが一文に集中することは滅多にないのでは」との指摘があった。そこで、2~3つの文をまとめて一内容と捉えることができるよう、カテゴリの名称を「一内容を踏まえる」と改め、カテゴリの定義を「各選択枝のキーセンテンスが隣接する2~3文に集中している」に変更した。

本調査

目的

予備調査で再定義・名称変更したキーセンテンスのカテゴリを用いて設問を操作し、キーセンテンス

の内容や文章中の範囲が錯乱枝の選ばれ方にどのような影響を及ぼすのか、能力群別に検討することであった。

方法

受検者 大学生 299名(男性 127名, 女性 172名)であった。

材料 予備調査と同様、私立大学で過去に出題された英語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用いた。ただし、設問・選択枝は本研究の目的に合うよう作成するため、英語の文章のみを引用した。

テスト冊子とテストデザイン テスト冊子は大問2つから構成され、第1問はすべての冊子に割り当てられた受検者が共通に解答する設問(3項目)、第2問は実験操作を行った設問(4項目)とした。第2問は冊子ごとに文章・設問・選択枝が異なっており、第2問の設問にはキーセンテンスの内容(5)×キーセンテンスの範囲(4)の20条件のうち1条件が割り当てられた。それぞれの条件が異なるテスト冊子で計3回含まれるよう配慮しながら、計15種類のテスト冊子を作成した。

各設問における選択枝は、寺尾・安永・石井・野口(2015)で特徴が見られた錯乱枝(否定語錯乱枝, 対義語錯乱枝, 因果関係の取り違い錯乱枝)と正答選択枝の4枝選択とした。

手続き 受検者には、テスト冊子が研究室で厳重に管理されること、研究への参加が任意であること、いつでも研究への参加を辞退できることなどを伝えた上で、同意を確認し、テストへの解答を求めた。テストへの解答時間は30分間であった。

結果

項目分析 受検者の項目反応パタン行列を、正答=1, 誤答=0とする2値反応パタン行列に変換し、項目困難度および項目識別力を求めた。項目困難度には正答率, 項目識別力にはI-T相関係数(当該

項目への反応と当該項目削除得点との相関係数)を用いた。項目分析の結果、極端に易しい項目(正答率 > .90)あるいは極端に難しい項目(正答率 < .10)は見られなかったが、項目識別力において負の値になるものが見られた。

IRTによる θ の推定 第1問・第2問をあわせて2値反応パターン行列に対し、1PLMを用いて受検者の潜在特性値 θ をMAP推定(ベイズモデル推定)した。事前分布を $N(0,1)$ として推定を行った。

受検者の群分け $\hat{\theta}$ を用いて受検者を能力低群・中群・高群の3群に分けた。古典的テスト理論の方法(池田, 1972)にならひ、 $\hat{\theta}$ の分布の下側27%を能力低群、中側46%を能力中群、上側27%を能力高群とした。

多項ロジスティック回帰分析 条件ごとに錯乱枝の選ばれ方を検討するため、多項ロジスティック回帰分析を行った。Model 1では、説明変数に能力群(3)、キーセンテンスの内容(5)、文章中のキーセンテンスの範囲(4)、能力群×キーセンテンスの内容、能力群×キーセンテンスの範囲、キーセンテンスの内容×キーセンテンスの範囲、能力群×キーセンテンスの内容×キーセンテンスの範囲を投入し、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対数オッズを予測した。その結果、推定値の計算結果が収束せず、推定値のSEにおいて不適解が検出されたため、Model 1は採用しなかった。

Model 2では、Model 1から2次の交互作用項(能力群×キーセンテンスの内容×キーセンテンスの範囲)を取り除いたモデルを立てたところ、推定値の計算結果が収束し、不適解も見られなかった。それぞれの主効果および交互作用を検討するため、尤度比カイ二乗検定を行った(Table 2)。

Table 2 Model 2の尤度比カイ二乗検定の結果

説明変数	LR- χ^2
能力群	207.70***
内容	30.05**
範囲	48.36***
能力群×内容	27.35
能力群×範囲	27.18†
内容×範囲	96.35***

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

デビアンズ残差分析 Model 2で交互作用項の効果が見られたものについて、標準化デビアンズ残差を用いた残差分析を行った。標準化デビアンズ残差は漸近的に $N(0,1)$ となることが知られており(Agresti, 2013; Bilder & Loughin, 2014)、標準化デビアンズ残差の絶対値が1.96を上回れば予測値と

観測値が離れていると判断する。本研究は、どの条件において錯乱枝の選択率が高くなっているか検討するため、以下では標準化デビアンズ残差が1.96を上回っている条件を取り上げ、このことをもって選択率が高くなっていると判断する。

まず、能力群×キーセンテンスの範囲の交互作用に関する検討を行った(Table 3)。

Table 3 能力群×文章中の範囲の残差分析の結果

	否定語	対義語	因果関係	正答
能力低群				
一内容	-1.354	1.784	-0.435	0.005
一段落	0.000	0.126	6.608	-6.735
複数段落	-0.659	0.598	-7.037	7.098
文章全体	2.004	-2.504	0.863	-0.363
能力中群				
一内容	-0.672	-0.398	5.047	-3.976
一段落	1.404	0.462	-6.552	4.686
複数段落	-1.188	0.176	4.690	-3.678
文章全体	0.451	-0.249	-3.181	2.979
能力高群				
一内容	2.026	-1.387	-4.613	3.974
一段落	-1.411	-0.590	-0.051	2.053
複数段落	1.844	-0.775	2.340	-3.409
文章全体	-2.457	2.750	2.321	-2.615

否定語錯乱枝について、能力低群ではキーセンテンスが文章全体に散らばっている条件で標準化デビアンズ残差が1.96を上回り、選択率が予測値よりも高くなっていた一方で、能力高群ではキーセンテンスが一内容を指していた条件で標準化デビアンズ残差が1.96を上回り、選択率が高くなっていた。

また、対義語錯乱枝については、能力高群のみで効果が見られ、キーセンテンスが文章全体に散らばっている条件で標準化デビアンズ残差が1.96を上回り、選択率が高くなっていた。

因果関係の取り違え錯乱枝について、能力低群ではキーセンテンスが同一段落中にある条件で標準化デビアンズ残差が1.96を上回り、選択率が高くなっていた一方、能力中群ではキーセンテンスが一内容を指す条件、複数段落にまたがる条件で標準化デビアンズ残差が1.96を上回り、選択率が高くなっていた。能力高群ではキーセンテンスが複数段落にまたがる条件と、文章全体に散らばっている条件で標準化デビアンズ残差が1.96を上回っており、選択率が高かった。

次に、キーセンテンスの内容×キーセンテンスの範囲の交互作用に関する検討を行った(Table 4)。

Table 4 内容 × 文章中の範囲に関する残差分析の結果

	否定語	対義語	因果関係	正答
事実				
一内容	-0.787	1.767	2.780	-3.761
一段落	-0.402	-1.738	-6.403	8.544
複数段落	0.066	-2.928	0.122	2.739
文章全体	1.122	2.903	3.505	-7.531
概念				
一内容	4.425	-2.565	-1.845	-0.015
一段落	-0.536	5.797	-1.569	-3.692
複数段落	-1.818	-2.405	2.205	2.199
文章全体	-2.071	-0.826	1.393	1.504
手続き・過程・背景				
一内容	-4.225	1.057	2.859	0.308
一段落	-0.857	-1.145	-2.786	4.788
複数段落	2.365	-4.131	4.853	-3.087
文章全体	2.719	4.214	-4.920	-2.012
原理				
一内容	0.051	-0.849	-6.589	7.387
一段落	-1.346	-0.691	5.553	-3.516
複数段落	3.590	10.787	-5.232	-9.145
文章全体	-2.298	-9.256	6.275	5.280
意見				
一内容	0.533	0.588	2.806	-3.927
一段落	3.132	-2.222	5.204	-6.113
複数段落	-4.191	-1.325	-1.765	7.281
文章全体	0.525	2.961	-6.242	2.756

否定語錯乱枝では、概念に関するキーセンテンスが一内容に集中している条件、手続き・過程・背景に関するキーセンテンスが複数段落にまたがっている条件や文章全体に散らばっている条件、原理に関するキーセンテンスが複数段落にまたがっている条件、意見に関するキーセンテンスが同一段落にある条件で標準化デビアンズ残差が 1.96 を上回り、選択率が予測値よりも高かった。

対義語錯乱枝では、事実、手続き・過程・背景、意見に関するキーセンテンスが文章全体に散らばっていた条件で、標準化デビアンズ残差が 1.96 を上回り、選択率が予測値よりも高くなっていった。一方、概念に関するキーセンテンスが同一段落にある条件、原理に関するキーセンテンスが複数段落にまたがって記述されていた条件で、標準化デビアンズ残差が 1.96 を超え、選択率が予測値よりも高くなっていった。

因果関係の取り違え錯乱枝では、事実に関するキーセンテンスが一内容に集中している条件、文章全体に散らばっている条件で標準化デビアンズ残差が 1.96 を上回っており、選択率が高かった。また、手続き・過程・背景、原理に関するキーセンテンスが一内容に集中している条件、複数段落にまたがっ

ている条件や、原理に関するキーセンテンスが同一段落にある条件、文章全体に散らばっている条件で標準化デビアンズ残差が 1.96 を超え、選択率が高くなっていった。

総合考察

本研究の目的は、英語文章読解テストにおける文章中のキーセンテンスの種類が、能力別にみた錯乱枝の選ばれ方に影響を及ぼすかどうか検討することであった。その結果、能力群 × キーセンテンスの範囲の交互作用が検出され、否定語錯乱枝の選ばれ方と対義語錯乱枝・因果関係錯乱枝の選ばれ方に大きな違いが見られた。否定語錯乱枝は、キーセンテンスが文章全体に散らばっていた場合に能力低群の選択率が高く、キーセンテンスが隣接していた場合に能力高群の選択率が高いという選ばれ方であった。一方、対義語錯乱枝や因果関係錯乱枝は、キーセンテンスが隣接していた場合に能力低群の選択率が高く、キーセンテンスが文章全体に散らばっていた場合に能力高群の選択率が高いという選ばれ方であった。錯乱枝の選ばれ方に対するキーセンテンスの範囲の効果が受検者の能力に応じて異なることを示した研究はこれまでに見られず、多枝選択式の英語文章読解問題を効果的に作成する一助となりうる。一方で、なぜそのような選ばれ方になるのかについては、今後検討が必要である。

近年、欧米では、受検者の認知過程に基づいた問題作成を行う必要性を主張する研究が増えてきた (e.g., Gorin, 2005). Embretson and Wetzel (1987) による英語文章読解問題の情報処理モデルでは、解答の決定に際し、選択枝に述べられている情報をエンコーディングし、文章中のキーセンテンスの表現と選択枝中の表現を照合した上で、選択枝の正しさを評価するというプロセスが提案されている。Embretson ら (1987) のモデルにおける解答決定のプロセスになぞらえて言えば、本研究はキーセンテンスと選択枝との照合を実験的に操作し、選択枝の正しさに関する評価について検討を行った研究であると言える。

本研究から得られる示唆として、英語文章読解テストにおける錯乱枝を作成する際、キーセンテンスの種類に留意する必要性が挙げられる。本研究の知見を生かし、今後の問題作成では、キーセンテンスの内容や文章中の範囲に配慮しながら作成することで、より効果的な錯乱枝となる可能性が考えられる。

[E-mail: terao.takahiro@a.mbox.nagoya-u.ac.jp]