

# 日本語文型を基にした多肢選択空欄問題の作成

植木 康之 齋藤博昭  
慶應義塾大学 理工学部  
{ueki,hxs}@nak.ics.keio.ac.jp

## 1 はじめに

多肢選択空欄問題とは、文章の一部が空欄となっている質問文に対し、その空欄に当てはまるのに適していると思われるものを選択肢の中から選ばせる問題のことである。選択肢は正解1つと、解答者を惑わす複数の不正解から構成されている。以下に例を示す。

- 2007年の干支は（ ）である。
  1. 申
  2. 酉
  3. 戌
  4. 亥

この多肢選択空欄問題は、今日世界中で評価テストを行う際に用いられている問題形式であり、学生の達成度を測るために効果的とされている。日本でも学力試験や資格試験などでこの形式の問題が取り入れられている。最近では、塾や家庭学習など学生が学校以外で勉強する機会が増えてきている。こうした場での学習効果を確認するために問題を作成し提供する機会は今後も増えていくだろう。

しかし多肢選択空欄問題の作成は依然として、人間が空欄とする項目を決めたり、不正解となる選択肢を考えることを手作業で行っていて、時間や労力を多く費やしているのが現状である。また、知識不足で答に自信のない解答者を惑わして間違えた答を選ばせるうまい不正解を考え出すことは、多肢選択問題を作成する上で難しい点とされる。

そこで本研究では、日本語の電子文書から手作業より効率的で質の良い多肢選択空欄問題を自動的に作成する手法を提案し、その評価を行う。

## 2 関連研究

コンピュータによるテスト問題の自動作成は比較的新しい研究分野の1つである。多肢選択問題については、英語のコーパスから自動的に作成す

る手法が提案されている [1]。Mitkov らは出現頻度を用いてコーパス中の重要語を選定してそれを正解としているまた彼らは WordNet を用いて上位語と等位語を検索し不正解の候補としている。さらには、作成された問題を人間が修正しやすいようにインタフェース環境を実装している [2]。多肢選択空欄問題の作成については、非母国語話者の英語学習のために英語の文の動詞を空欄とした問題を作成する手法が提案されている [3]。

本研究では日本語のコーパスから多肢選択空欄問題を自動的に作成する。本手法は文中のある特定の語を説明している文構造に着目し、そこから正解となる語を決定している。また本手法では日本語文書を入力とするので等位語の検索に EDR 辞書の概念辞書 [5] を使用している。

## 3 問題作成手法

問題作成にあたって、本研究の提案手法は以下の前提に基づいている。

一つは、質問文の空欄部となるものは空欄部以外の文章を読み取ることで一意に特定されなければならないということである。そこで本研究では、文の内容で一意に特定することができる名詞や名詞句が含まれていると考えられる文構造を考え、それを基にしたルールベースによる質問文の選定を行っている。例えば、「ブラックホールは太陽のように光を発している恒星のうち質量が太陽の数倍を超えるような重い星が超新星爆発を起こし、一生を終えた姿とされる。」という文は、「ブラックホール」について説明している文であり、「ブラックホール」の部分を変換して多肢選択空欄問題の質問文とすることができる。

さらに本研究では、正解の意味に近く、なおかつ余計なヒントを与えない不正解が解答者を惑わす良い選択肢となることを前提としている。例えば、「ブラックホール」という正解に対しては、「ホワイトホール」や「ワームホール」といったものが

不正解として適切であり、「タイムマシン」や「掃除機」などは相応しくないと考えられる。

以上の前提から、日本語コーパスから多肢選択空欄問題を作成するシステムを実装した。システムは質問文生成と不正解決定という2つのタスクを実行するよう構成されている。作成される問題形式は、一続きの1~3文からなる質問文と正解を1つ含む名詞4つの選択肢となっている。

### 3.1 質問文生成

本研究ではまず、入力文書から質問文となりうる文を選定するために、入力文書の各文の形態素解析を行う。形態素解析には茶筌 [4] を用いる。

次に、ある特定の文型からなる文構造ルールと合致している文を検索し、その文の正解となる名詞（以下、正解語）の部分为空欄化したものを問題の質問文とする。文構造ルールは、それぞれ文の前半部に現れた形式に応じて述語形式が定められていて、述語動詞の活用形に関係なくこれらの形式の文である場合に合致とする。正解語は、前半部もしくは述語末尾の直前に置かれている名詞<sup>1</sup>とし、両方に名詞が置かれている場合にどちらを優先してとるかは前半部と述語の組み合わせによって異なる。「名詞 A + 『の』 + 名詞 B」というように助詞の「の」もしくは「という」が2つ以上の名詞を連ねている場合は、先頭の名詞 A を正解語とする。前半部と述語の両方で末尾の直前に名詞が出現していない文は不一致とする。文構造ルール一覧を表 1 に示す。

また、質問文中に「この」などの指示語が含まれている場合は指示語内容を補完するために直前の文も質問文として追加する。直前の文にも指示語が含まれている場合はさらに直前の文も追加し、質問文は最大で3文までの構成とする。但し、正解語が「これ」や「それ」などの指示語である場合その質問文は却下される。さらに、正解語が直前の文に含まれている場合も質問文として追加する。この場合でも質問文は最大で3文までの構成とする。

例えば、「これに対して、ヒトの卵子は体外での操作に強いとされる。」という文は文構造ルール「~は~とされる」に一致する。正解語の優先部は「ヒトの卵子」となり、これは「A の B」という形なので正解語は「ヒト」となる。さらに「こ

<sup>1</sup>形態素解析によって「未知語」と判定された語も含む。また、非自立語である「の」や「という」という名詞はこれに含まれない。

表 1: 質問文選定のための文構造ルール一覧

前半部形式	述語形式	優先部
~は, ~とは	~である	前半部
〃	~となる	前半部
〃	~とされる	前半部
〃	~といわれる	前半部
〃	~とよばれる	述語
〃	~を意味する	前半部
~を	~という	述語
〃	~とよぶ	述語
~には	~がある	前半部
~として	~がある	述語

れ」という指示語が含まれているのでこの直前の文も含んだ質問文が生成される。

### 3.2 不正解決定

まず、EDR 辞書の単語辞書と概念辞書 [5] を用いて、正解語に意味が近く不正解の選択肢となるのに相応しい概念を検索する。概念辞書には対応する上位概念と下位概念の組合せが ID によって羅列してあり、単語辞書は単語とそれに対応する ID を検索するのに用いる。この概念辞書を使って正解語と同じ上位概念を持つ等位語を検索し、それらを不正解の候補（以下、不正解語）とする。

もし検索された不正解語が2つ以下で、なおかつ正解語が2つ以上の名詞が結びついてきた複合名詞であるならば、正解語の末端の単名詞を除いたものの等位語を検索し、見つければ除かれていた単名詞を等位語の末端に付け戻して不正解語とする。さらに見つからない場合は同様の操作を行い、等位語が見つかるかもしくは検索する語が単名詞となるまで繰り返す。例えば、「核燃料サイクル」という正解語は等位語を検索しても見つからないので、次に正解語の末端の単名詞「サイクル」を除いて「核燃料」で検索する。すると「光圧」や「運動エネルギー」等3つ以上の等位語が検索される。最後にそれらの等位語の末尾に「サイクル」という名詞を結びつけて不正解語とする。最終的に不正解語が2つ以下だった場合は選択肢に不正解は出力されない。

正解語が数字もしくは数字のあとに名詞が結びついた複合名詞である場合は、数字の最も高いケタを1から9でランダムに変えたものを不正解語

とする。例えば、「300メートル」という正解語に対しては、「200メートル」や「500メートル」という不正解語が出力される。

次に、選択肢の候補となる多数の不正解語から3つの語を選定し、実際に問題の選択肢として出力する。選定はランダムに行うが、不正解語が10個以上ある場合は入力文書の出現頻度を測定し、頻度の高い10個の不正解に絞り込んでからランダムに選定する。これは、同じ上位概念を持つ別々の正解が現れた場合に、共通な不正解が出力しないようにするためである。

## 4 実験

本研究の問題作成手法の効率性と作成された問題の質を調べるために評価実験を行った。

問題を評価するにあたり、システムが作成した問題についてはあらかじめ質問文と不正解に関する評価基準を設け、問題を評価ごとに分ける工程を行った。質問文に関しての評価基準で4段階の評価をし、その中で「適切」の評価となった問題のみさらに3段階に評価した。評価基準を表2に示す。

表 2: システムによる作成問題の評価基準

適切	問題内容については修正の必要なし	
	A	何も修正の必要がないもの
	B	余計な接続詞の削除など質問文中で細かい修正の必要があるもの
	C	不正解が出力されていないもの
可1	質問文はそのまま空欄位置（正解語）を別のものに必要があるもの	
可2	正解語についての説明不足で添削等の修正の必要があるもの	
不可	質問文全体の内容について問題にする価値がないもの	

### 4.1 作成手法の効率性

システムと人間がそれぞれ98年度毎日新聞[6]科学面全文記事から多肢選択空欄問題を作成し、それぞれの作成時間と作成問題数から1問あたりの平均作成時間を測定した。

システムの作成時間はプログラム処理の時間に評価時間と修正時間を足した値とした。評価時間は1問につき30秒としシステムの全作成問題を対象とした。また修正時間は1問につき3分とし「不可」以外の評価を受けた問題を対象とした。さらにシステムの作成問題数は「不可」以外の評価を受け修正された問題の数とした。

人間の被験者は2人で、1人ずつ別々に時間を設け、その制限時間内にできるだけ多くの問題を作成してもらった。人間の作成時間と作成問題数は2人の合計とした。問題形式は1問につき、一続きの1~3文の質問文と正解が1つ含まれる名詞4つの選択肢に指定した。また質問文中に出現する正解語はすべて空欄にするよう指示した。

### 4.2 作成問題の質

効率性を調べる実験で作成された問題を採用したテストを実施し、被験者の正答率から問題の質を評価した。またテストと同時に紙面アンケートを行い、実際に問題を解いてみて思った問題内容の難易度について「簡単」、「適切」、「難解」の3段階で主観評価を行った。

被験者は14人で、効率性の評価実験の被験者は含まれていない。被験者には1問1分以内で解答してもらった。テストは全30問でその内人間が作成した問題を14問、システムが作成した問題を16問とした。システムが作成した問題は「適切」という評価を受けた32問からランダムに選び、必要があれば修正を行ったものとした。修正されたのはその内8問であった。

問題内容の難易度については、全回答数に対する「適切」の割合を評価した。また「簡単」~「難解」を1~3とする評価値をとり、システムと人間の場合で各被験者の評価値の平均を比較し、評価値が2に近く良評価である方の被験者数をそれぞれ数えて評価した。

### 4.3 結果

表 3: 作成手法の効率性についての評価結果

	作成時間	問題数	平均作成時間
システム	540分	112	4分49秒
人間	120分	14	8分34秒

表 4: 作成問題の質についての評価結果

	正答率			主観評価	
	平均	最高	最低	適切	良評
システム	70%	81%	50%	23%	8人
人間	61%	93%	29%	30%	2人

システムの出力全 323 問を評価分けした結果、全体の 34.7%にあたる 112 問が「不可」以外の評価となり価値のある問題となった。またその内の 25 問が修正の必要がないとされる「適切 A」の評価となりそれは全体の 7.7%であった。

作成手法の効率性についての評価結果を表 3 に示す。1 問あたりの平均作成時間は本研究手法によるシステムを用いた場合では 4 分 49 秒で、人間による手作業で作成した場合の 8 分 34 秒の 5～6 割の時間となった。

作成問題の質についての評価結果を表 4 に示す。平均正答率はシステムによる問題が 70%、人間による問題が 61%となり、若干システムを用いた方が簡単だったという結果がでた。システムによる問題の最高正答率が 81%、最低正答率が 50%だったのに対し、人間の場合はそれぞれ 93%、29%となり、システムに比べ大きな開きがあった。またアンケートによる主観評価では作成問題の内「適切」の評価を受けた問題数の割合は本研究手法による場合が 23%、人間による場合が 30%となった。さらに各被験者の評価値の平均を比較すると本研究手法の方を良評価とした人数は過半数の 8 人であった。全体の評価値の平均は本研究手法による場合が 2.4、人間による場合が 2.3 で若干システムによる問題の方が難しいという結果となった。

#### 4.4 考察

本研究手法による問題作成は人間による手作業と比べ効率性において向上がみられる。作成問題の質については、全体として難易度は少し簡単であるが、内容的には難しい印象を受けている。また主観評価で「適切」となった問題はやや少なかったものの、良評価とする被験者数が多く、最高正答率と最低正答率の差が小さいことから、確認テストやチェックテスト向きの問題であるといえる。

## 5 まとめ

多肢選択空欄問題の作成において、本研究手法により作成にかかる時間と労力を減らすことができた。また特定の用途において問題の質が向上した。

今後の課題には、さらに効率性を高めるために、係り受けや照応関係に対応した文構造ルールの追加・削除、異なる辞書の追加、人間が問題を修正しやすくなるインタフェースの実装が挙げられる。また評価基準による初期段階の評価についても人間に行ってもらい、実験規模を拡大することでさらに正確な評価結果を得ることが必要である。さらに教科書コーパスを用いた実験を行い、将来的に教育現場や e ラーニングへの応用が考えられる。

## 参考文献

- [1] Mitkov, R. and Li An Ha (2003) Computer-aided generation of multiple-choice tests. *In Proceedings of the First Workshop on Building Educational Applications using Natural Language Processing*, pp. 17-22.
- [2] Mitkov, R., Li An Ha, and Karamanis, N. (2006) A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering* 12 (2), pp. 177-194.
- [3] Sumita, E., Sugaya, F., and Yamamoto, S. (2005) Measuring Non-native Speakers' Proficiency of English by Using a Test with Automatically-Generated Fill-in-the-Blank Questions. *In Proceedings of the Second Workshop on Building Educational Applications Using Natural Language Processing*, pp. 61-68.
- [4] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: “形態素解析システム『茶筌』version 2.3.3 使用説明書, <http://chasen.naist.jp/hiki/ChaSen/>, 奈良先端科学技術大学院大学松本研究室”, (2003).
- [5] 日本電子化辞書研究所: “EDR 電子化辞書 1.5 版仕様説明書”, (1995)
- [6] 毎日新聞: “CD-毎日新聞 98 年版”.