

誤回答情報を用いた多肢選択形式作問支援システムに おける誤答選択肢の評価

菅原 遼介^{†1} 高木正則^{†2}

近年、情報処理技術者試験や TOEIC 等の資格試験作成現場では、作問における手間が大きいことが問題となっている。資格試験作成現場等では採点や管理のしやすさから多肢選択形式問題の作成需要が高く、多肢選択形式問題を採用する機会が多い。しかし、多肢選択形式問題の作成は問題文と正答に加え、誤答選択肢の作成が必要となるが、誤答選択肢作成には専門的な知識が必要となり、多くの負担が掛かるとされている。そこで、我々は誤答選択肢を自動生成することで作問負担を軽減させるシステムを提案してきた。既存の問題自動生成システムでは、web 上に存在する情報の整形を行うことで知識ベースの構築と問題の自動生成を行っている。また、単語間の類似性までを管理した知識ベースの構築には多くの手間がかかる。そこで我々は、知識ベース構築の手法として誤回答の活用を検討し、誤回答を活用した誤答選択肢自動生成システムの開発を行った。このシステムでは、誤回答情報の他に、過去問題中の類似問題、web 上の情報を用いて誤答選択肢自動生成を行うことでより精度の向上を目指している。本稿では誤回答情報を用いて誤答選択肢を生成した際の多肢選択形式問題の有用性の評価を行った。本研究では盛岡もの識り検定試験(もりけん)というご当地検定試験を対象にしている。実験の結果、誤回答データの誤答選択肢への活用可能性が示唆された。

Evaluation of Mistake Choices in Quiz Creation Support System for Multiple-Choice Question Format with Mistake Answer Information

RYOUSUKE SUGAWARA^{†1} MASANORI TAKAGI^{†2}

In recent years, there is a growing problem for a lot of labor of quiz creation in test creation field such as information technology engineer examination and TOEIC. There is a growing need for creation of multiple-choice question in test creation field because the question format can be scored automatically and managed easily. Therefore, It's often the case that they go on to multiple-choice question. However, the creation of incorrect answer choices similar to a correct answer choice requires specialized experience and a lot of labor. In order to solve this problem, we had been proposing about quiz creation support system by automatic generation of the incorrect answers. Our related systems generate a quiz using knowledge base constructed using information on the web. Furthermore, the construction of knowledge base which manages the similarity between words requires a great deal of labor. Here we proposed a method of constructing knowledge base using wrong answers of a written question and developed an automatic generation system of incorrect answer choices using the knowledge base. The system want to be accuracy enhancement of automatically create mistake choices used mistake answer information and similar question in past question and information on the web in the future. In this paper, we evaluate validity of multiple-choice question from the system automatically create mistake choices used mistake answer information. This research is aimed at local knowledge exam called Morioka Monoshiri Kentei Shiken(Moriken). In result of trial experiment, we indicated the wrong answer can use in order to generate incorrect answer choices automatically.

1. はじめに

近年、情報処理技術者試験 [1]や TOICE [2]などの資格試験作成現場では、作問における手間が大きいことが問題となっている。多くの資格試験では出題に多肢選択形式問題を採用しているが、多肢選択形式問題の作成は問題文と正答に加え、誤答選択肢の作成が必要となるため、一問一答形式問題に比べて作問負担が大きい。試験採点時の手間を考えると、一問一答形式問題は回答を自由に記述出来るため、曖昧な表現や表記の揺れなどで機械的に採点を行うことが困難となり、採点の手間が大きくなる。一方で

多肢選択形式問題は、受検者がどの選択肢を選択しているかの情報で機械的に採点出来るため採点の負担は少ない。この事からも、多肢選択形式問題の作成ニーズが高いことが分かる。

この問題に対し我々は多肢選択形式の作問の中でも負担の大きい誤答選択肢の作成負担の軽減を目的とし、誤答選択肢自動生成機能を有した作問支援システムの構築を検討してきた [3]。

先行研究では、岩手県盛岡市で開催されているご当地検定試験”盛岡もの識り検定試験”(以下、もりけん)で出題された一問一答形式問題の誤回答を活用し、誤答選択肢の自動生成を行った。しかし、誤答選択肢を自動生成した問題が実際の試験で利用出来る程の質を有しているものか評価は充分には行われていない [4]。

^{†1} 岩手県立大学
Iwate Prefectural University

表 1. 出題形式の違いによる作問時間の差異

	一問一答	多肢選択	(誤答選択肢)
入力時間	1分20秒	1分51秒	(31秒)
思考時間	3分29秒	4分41秒	(1分12秒)
合計時間	4分49秒	6分32秒	(1分43秒)

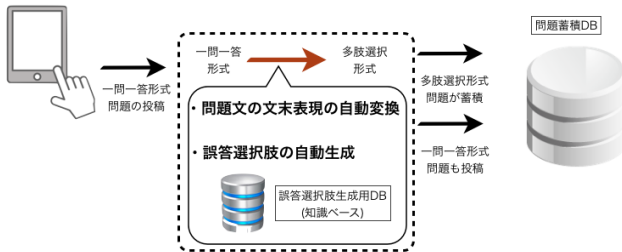


図 1. 問題形式自動変換機能の概要

本稿では、開発した作問支援システムについて述べたあと、その中で、一問一答形式問題の誤回答を活用した誤答選択肢自動生成手法の評価を行い、誤回答の活用可能性を考察する。

2. 目的達成へのアプローチ

2.1 作業時間軽減の見込み

前述したように、多肢選択形式問題の作問の場合、誤答選択肢の作成が大きな負担になる。学生 20 名に協力してもらい誤答選択肢の作成負担を調査したところ、一問一答形式問題と多肢選択形式問題では 1 問作成するに当たり 1 分 43 秒の差が生じることが分かった [5]。表 1 に出題形式の違いによる作問時間の差異を示す。もりけんの場合 1 年に作成する多肢選択形式問題は 230 問ある。誤答選択肢の自動生成ができれば 23690 分の作業時間が短縮できることが見込まれる。

2.2 出題形式変換による作問支援

2.1 節の内容も踏まえ、我々は作問負担の少ない一問一答形式問題をニーズの高い多肢選択形式問題に自動変換することによる作問支援の方法を考えた。本手法の概要を図 1 に示す。なお、本研究では、正答・誤答選択肢が名詞の問題を対象としている。

本手法では、作問者に問題文と正答を入力することで一問一答形式問題を作成してもらう。一問一答形式問題と多肢選択形式問題では問題文の文末表現が「～を書きなさい」「～はどれか」といったように異なるため、問題文の文末表現の変換が必要になる。その問題文の文末表現を自動変換したあと、生成した誤答選択肢を設定して多肢選択形式へと変換する。

正答選択肢と誤答選択肢との類似度は問題の難易度に影響を与えるため、正答との類似度を考慮した誤答選択肢の選定が求められる。以上より、誤答選択肢の自動生成のた

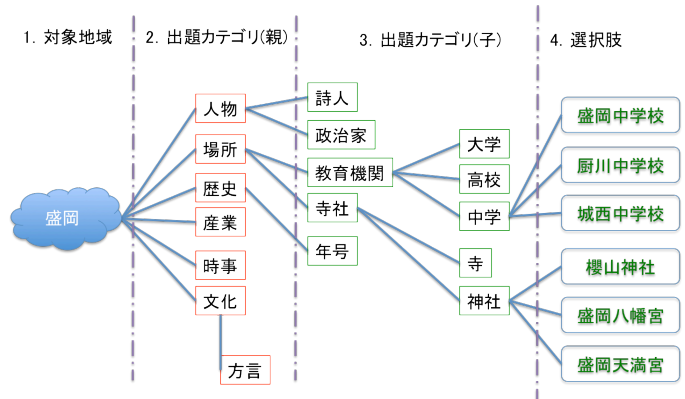


図 2. 単語間の類似度を考慮した知識ベースの例

めには、

- (1)単語間の類似度が管理された知識ベースの構築
 - (2)知識ベースからの単語（誤答選択肢）の選出
- の課題解決が必要となる。

2.3 知識ベースの構築と対象試験の特徴

前述した課題を解決するためには、単語間の概念距離を算出した知識ベースを用意する必要がある。しかし、これらを手作業で実施するには手間が大きいため自動化が望まれる。必要となる知識ベースの例を図 2 に示す。

知識ベースの構築を自動化し、作問を支援するシステムは様々研究されているが、これらのシステムでは、web 上に存在する情報を整形することで作問支援を行っている [6] [7] [8]。この方法は、情報処理技術者試験や TOEIC 等を対象としており、作問を行うための知識が web 上で網羅的に管理された知識ベースが構築されている場合に活用できる。しかし、新技術やイベント情報、あまり話題になっていない知識等も出題されるご当地検定の場合、情報が web 上に存在しないことも多く、このような問題を自動生成する際には対応が難しい。以上のことからご当地検定試験に対応した問題を自動生成するシステムを現状の手法で構築することは困難であると言える。

3. 誤答選択肢自動生成機能の提案

作問者によって作成される誤答選択肢と同程度の信頼性をもつ誤答選択肢の自動生成を目的として、本研究では 3 つの誤答選択肢自動生成手法を提案する。

3.1 誤回答を利用した誤答選択肢の生成

受検者が一問一答形式問題を回答した際の誤回答を収集・整理し、多肢選択形式問題の誤答選択肢に利用する

(1)誤回答の管理方法

一問一答形式問題に回答した際に得られる誤回答は正答と何らかの関連を持っていると考えられる。例えば、有料の資格試験や大学の入学試験などでは、答えが分からない問題であっても、受検者が持っている知識の中から憶測で答えを記入することや、正解だと確信していた答えであっ

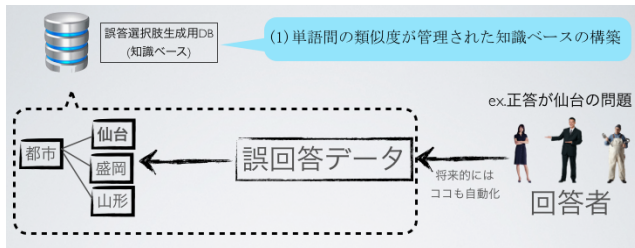


図 3. 誤回答収集のイメージ

でも不正解になってしまう場合もある。このような誤回答情報には今後の試験問題の作成等に役立つ有益な情報が含まれていると考えられるが、これらを誤答選択肢として再利用する研究は著者が知る限り存在しない。誤回答収集のイメージを図 3 に示す。例えば、仙台が答えとなる問題(例:牛タンはどこの名物でしょう)に対しては、山形、秋田、盛岡など仙台に近い都市名を誤って回答する可能性が考えられる。この場合、仙台、山形、秋田、盛岡は何かしらの関連のある単語として扱い、データベースに蓄積する。このように、正答と誤答を関連付けてデータベースで管理することによって、データベースの中のある単語が正答に設定されたら、その言葉と関連付けられている言葉を誤答選択肢として自動的に設定できるようになる。例えば、盛岡が答えとなる問題(例:わんこそばはどこの名物でしょう)の場合は、仙台、山形、秋田を誤答選択肢として自動的に設定することが可能となる。

(2) 誤答選択肢生成のアルゴリズム

誤回答を用いた誤答選択肢生成の流れを図 4 に示す。誤答選択肢には、答えが同じ一問一答形式問題の誤回答が多いものから優先的に選択する。図 4 の場合、舟場橋という正答に対して繫大橋と答えた人が 6 人で最も多く、太田橋と答えた人が次に多かったということになる。また、繫大橋が 3 名、太田橋が 3 名、船場橋が 2 名、船田橋が 2 名のように誤回答のあった回数が同数であった場合は同じ誤回答数の単語からランダムで選出すこととした。表 2 に誤答選択肢が自動生成された問題の例を示す。

3.2 類似問題を利用した誤答選択肢の生成

もりけんは平成 18 年から試験が実施されており、既に 1000 を超える問題が作成されている。この過去問題中の多肢選択形式問題に含まれる選択肢を新たに作成する問題の誤答選択肢として活用する。図 5 に類似問題を利用した誤答選択肢の生成手順を示す。

表 2. 自動生成された問題の一部

NO	問題文	選択肢a	選択肢b	選択肢c	選択肢d(正解)
48	今年、岩手県単独として 32 年ぶりに「いわてデスティネーションキャンペーン」が実施されました。前回のキャッチフレーズは「詩情豊かな岩手路」でしたが、今回のキャッチコピーは何でしたか。	復興負けない岩手	そういう旅に私はしたい	黄金の国いわて	イートロープいわて物語~そういう旅に私はしたい。
49	金山の里・大ヶ生にある自然散策コースで、志波城から見ると冬至、徳丹城から見ると夏至の太陽が、その山頂から昇るといわれている山の名前は次のうちどれか。	飯岡山	早池峰山	浅島山	朝島山
50	このしの 10 月に開催された「もりおか映画祭 2012」で、盛岡が舞台となった作品が1本上映されました。その映画のタイトルは次のうちどれか。	ハル	馬	るろうに剣心	(ハル)

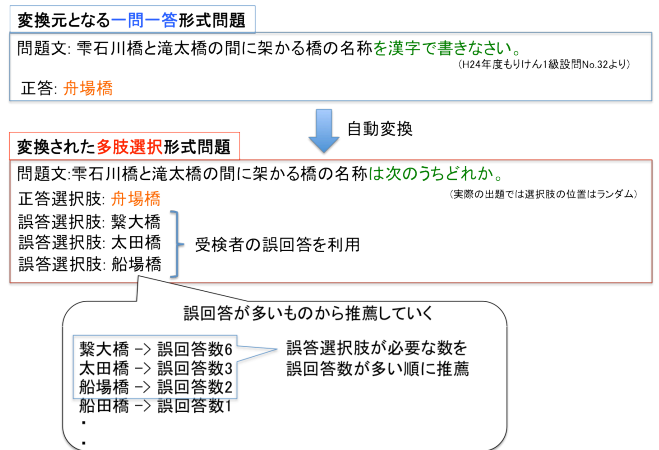


図 4. 誤回答を用いた誤答選択肢設定の流れ

(1) 一問一答形式問題からキーワードを抽出

ユーザが作問する一問一答形式問題の情報を形態素解析し、TF-IDF を用いてキーワードを抽出する。形態素解析には Yahoo!JAPAN の提供する日本語形態素解析 API を利用した。ここで入力する情報は、もりけんが出題された全問題を解答できる学習支援システム「もりけん web クイズ」 [9] に登録されている問題文・解答・タグとした。

(2) 過去問題から類似問題を検索

過去のもりけんが出題された多肢選択形式問題の問題文・選択肢・答え・解説に対して(1) 同様に形態素解析と TF-IDF を用いてキーワードを抽出する。(1)で抽出されたキーワードと共通するキーワードを含む過去問題を類似問題として選出する。

(2) 選択肢の生成

ユーザが入力した一問一答形式問題の解答を正解選択肢として設定し、類似問題中の選択肢の中から、正解選択肢と異なる選択肢をランダムに選出し、誤答選択肢に設定する。

3.3 Web 上の情報を活用した誤答選択肢の自動生成

誤答選択肢の生成に活用できる可能性のある Web 上の情報として、日本語 Wikipedia オントロジー (以下、JWO) [10]

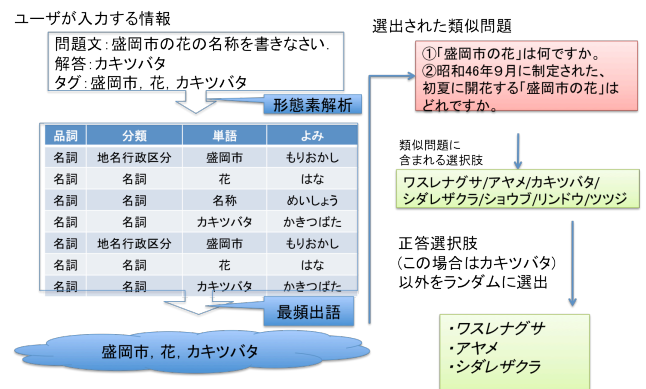


図 5. 過去問題からの誤答選択肢設定の流れ 2

と DBpedia [11]がある。JWO はデータが公開されていて利用が簡易であり、DBpedia よりも日本語の固有知識に対する情報量が多い。そこで、本研究では JWO を用いて誤答選択肢の生成を検討した。

3.3.1 日本語 Wikipedia オントロジー検索(JWO Search)

JWO は web 上で検索・閲覧できる日本語 Wikipedia オントロジー検索サービスを公開している。Wikipedia 上に投稿された記事の最初の段落を概要(アブストラクト)、Wikipedia 上の一覧記事を上位概念(クラス)、Wikipedia のリダイレクトリンクを用いて、同義語の抽出などを行い、Wikipedia からオントロジー構築を行っている。図 6 に JWO 検索による岩手県立大学の検索結果ページを示す。

3.3.2 JWO を用いた誤答選択肢自動生成

入力された問題文と正答単語から JWO を用いて誤答選択肢を自動生成する流れを以下に示す。

- ① 正答単語の表記揺れ補正
- ② 上位概念の抽出
- ③ 上位概念の選定
- ④ 単語の選定
- ⑤ 単語の補正

① 正答単語の表記揺れ補正

入力される正答単語が類義語や関連語、表記揺れ等で同じ意味を持つが違う表記になっている場合、それらを統一しないと概念の捕捉が困難になる。そこで Google 検索と wikipedia の情報を用いて表記揺れの補正を行う。図 7 に単語の表記揺れ補正の流れを示す。Google 検索では検索ワードに対して、検索結果の文章が取得出来るが、検索結果

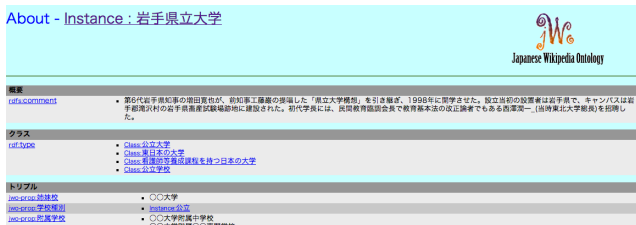


図 6. 岩手県立大学の JWO 検索結果ページ

には検索ワードが含まれるため、検索結果が特徴語として選出される。また Wikipedia のページが検索された場合は Wikipedia の情報を優先的に取得することとした。Wikipedia は複数記述が存在する同義語に関してはリダイレクトを行い1つのページに統合する機能があるため、この機能を活用して表記揺れを補正する。

② 上位概念の抽出

①で補正された単語に対し、JWO で検索を行う。JWO では Wikipedia の一覧記事に含まれる単語をクラス [rdf:type] とし、記事名を下位語とする語彙を上位語 [rdf:hyper]として定義している。単語のクラスと上位語の情報から上位概念を抽出する。図 8 に上位概念の抽出例を示す。

③ 上位概念の選定

入力された問題文に対して形態素解析を行い、名詞情報を抽出する。ここで抽出された名詞情報と②で抽出された概念のマッチングを行い、概念が一致したものを上位概念として選定する。図 9 に概念選定の流れを示す。この例では「慶應義塾の創設者として有名な教育者の名前を答えよ。」という問題文を形態素解析し、慶應義塾/創設者/有名/教育者/名前という名詞が抽出される。この名詞の中から②で抽出された上位概念と一致する名詞を検索し、図 9 では「教育者」という概念が誤答選択肢の上位概念となる。

日本語wikipediaオントロジーを利用

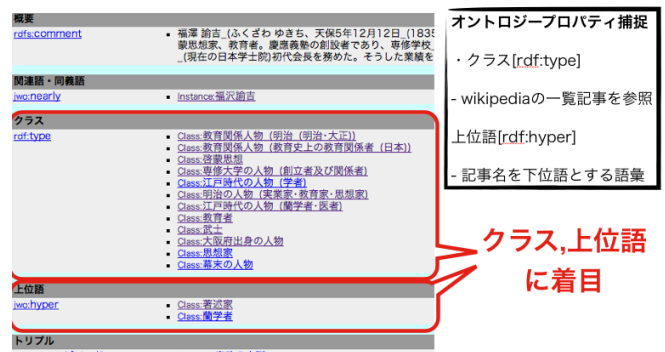


図 8. 上位概念の抽出例

Google検索(wikipedia)とTf-idfを利用

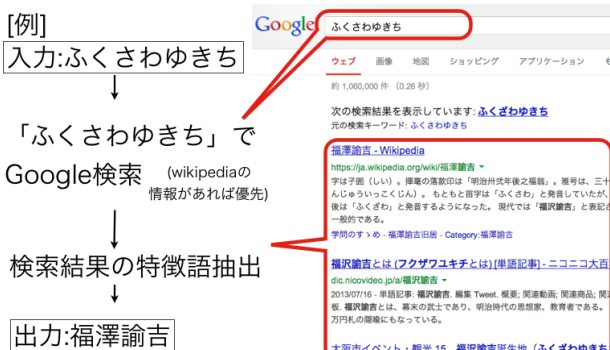


図 7. Google と Wikipedia を用いた単語補正の流れ

問題文を形態素解析した結果を利用

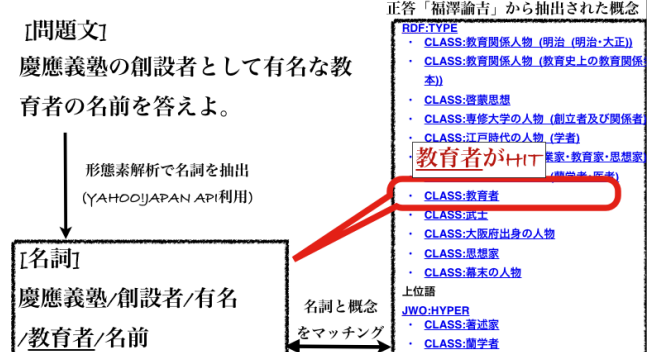


図 9. 利用する上位概念の選定

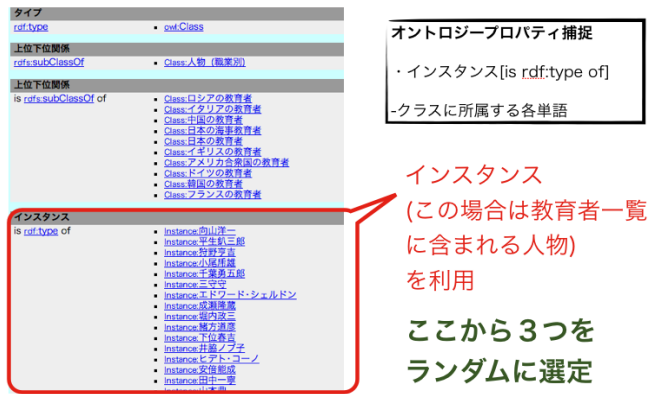


図 10. 誤答選択肢となる単語の選定

④単語の選定

③で選定された上位概念に属する単語で、正答単語以外の単語をランダムに3つ推薦する。単語選定のイメージを図 11 に示す。③で上位概念が教育者となったので、この場合は教育者を上位概念とする人物名一覧を取得し、その中に含まれる人物を誤答選択肢として設定する。

⑤単語の補正

福澤諭吉が正答単語となる問題において、図 8~10 の手順で誤答選択肢を生成した場合、「阿部熊成、向山洋一、エドワード・シェルドン、福澤諭吉」という選択肢が生成される可能性がある。ここでエドワード・シェルドンのみ日本人ではなく、文字列が長くなる。これは1つだけ属性(ここでは出身国)の異なる選択肢となっており、このような選択肢は望ましくない。このような場合は、④で選定された単語について、レーベンシュタイン距離(編集距離) [12]を用いて文字列類似度の補正を行う。図 11 に誤答選択肢の補正イメージを示す。正答単語を基準としてレーベンシュタイン距離(文字列の編集距離)の大きさを補正し、③で選定された他の上位概念に置換することで「阿部熊成/向山洋一/狩野亨吉/福澤諭吉」といった選択肢に補正する。この例では福澤諭吉を基準にすると阿部熊成、向山洋一は距離が12、エドワード・シェルドンは32となり、距離が大きい(文字列が似ていない)ため、距離が近い候補に置き換えている。

レーベンシュタイン距離(編集距離)を利用

選定された選択肢		補正された選択肢	
人物名	レーベンシュタイン距離 (福澤諭吉を基準)	人物名	レーベンシュタイン距離 (福澤諭吉を基準)
阿部熊成	12	阿部熊成	12
向山洋一	12	向山洋一	12
エドワード・シェルドン	32	狩野亨吉	8
福澤諭吉	(正答なので固定)	福澤諭吉	(正答なので固定)

【出力例】

問題文:
 慶應義塾の創設者として有名な教育者の名前を選べ。

誤答選択肢:
 阿部熊成、向山洋一、狩野亨吉

正答:
 福澤諭吉

図 11. レーベンシュタイン距離による誤答選択肢補正

4. テスト作問支援システムの開発

4.1 システムの全体概要

3章で述べた提案内容をもとに、多肢選択形式作問支援機能を有するテスト作成システム KenteiMaker(ケンテイメーカー)の開発を行った。KenteiMaker は、もりけん作問委員会が実際に作問作業を行う際に作業負担を軽減させる目的で開発・機能追加を行っている。将来的に汎用的な試験作成現場での利用を考慮し、試験毎に作問者グループを作り問題を共有・管理できるようになっている。作問グループのトップページを図 12 に示す。グループで作問作業をする際のタスクの進捗状況が分かりやすいように、グループでの作問達成度や個人毎の作問数を表示する機能を提供している。

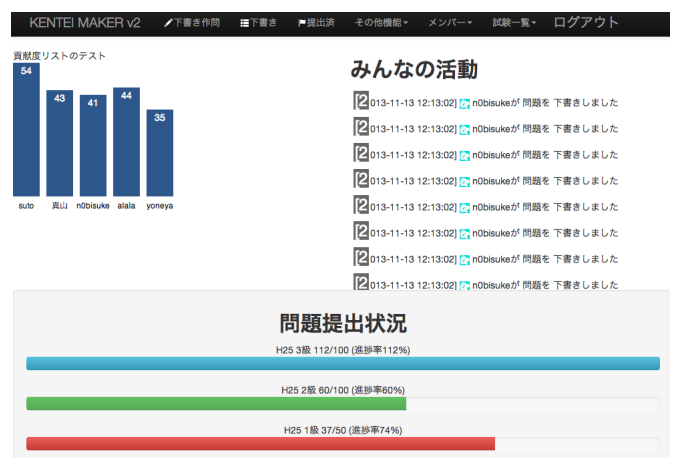


図 12. グループトップ画面

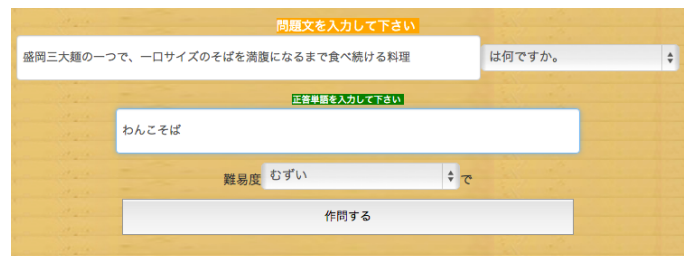


図 13 誤答選択肢自動生成入力画面

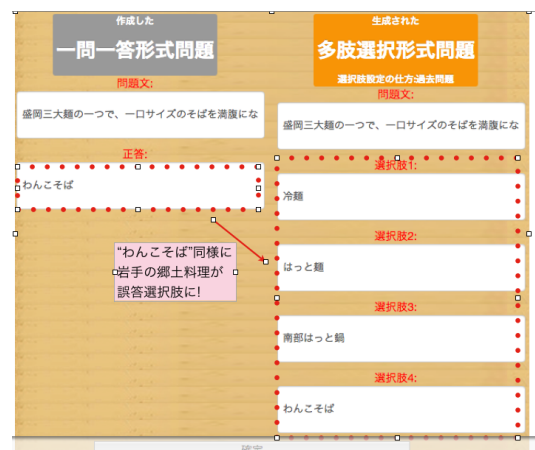


図 14. 誤答選択肢自動設定

4.1.1 誤答選択肢自動生成機能

誤答選択肢を生成したい一問一答形式の問題をシステムに入力する。入力画面を図 13 に示す。一問一答形式の問題文は文末表現をあらかじめ用意しているテンプレートで固定する。これにより、多肢選択形式問題の文末表現への自動変換を可能にしている。例えば、一問一答形式問題の文末表現として、「～は何ですか」「～を書きなさい」などのテンプレート表現を用意し、この部分を「～はどれですか」「～を選びなさい」などの文末表現に変換することで、多肢選択形式問題の問題文へと変換する。

図 14 に誤答選択肢自動生成の例を示す。図 13 で入力した内容をもとに、正答単語「わんこそば」の誤答選択肢として推薦している。この場合は「冷麺」、「はっと麺」、「南部はっと鍋」といったご当地グルメが誤答選択肢として設定されており、正答単語の「わんこそば」と同じ概念を持っていることが分かる。この例では、前述した誤回答、過去問題、JWO の情報を利用して生成している。

5. 誤回答を用いた誤答選択肢の評価

4 章で紹介した KenteiMaker の誤答選択肢自動設定機能を使うことで、一問一答形式の問題を多肢選択形式問題へと自動変換することが可能となったが、変換後の問題が試験等で利用出来る質を担保しているのか評価した。システムで誤答選択肢を生成した多肢選択形式が実際の試験で利用可能であることが証明できれば、本手法の実用化が期待できる。本稿では、現時点で行った誤回答を用いた誤答選択肢生成手法の評価実験について述べる。

5.1 評価指標

問題の質を判断する基準として、永岡らが行った研究 [13] を参考とした。永岡らはテスト作成力と学力との関連性を分析することを目的とし、大学生に対し、英語の学力およびテスト作成力の測定と、関連性の分析を行っている。

5.1.1 良質なテスト

永岡らは『テストは通常、学力の測定・評価を目的とする手段として作成され用いられるが、その目的に対して良質なものとそうではないものがある。特に、悪問・奇問と呼ばれる項目（問題）が多く存在する場合は良質なテストとは言えない。悪問・奇問には、解く人の認識により意味合いが異なる問題（悪定義問題）、問題解決のアルゴリズムの構築が難しい問題（悪構造問題）、また、学力の低い人が解けるのに高い人が解けないといった学力の測定として不十分である問題、などがあげられる。』 [14] と述べており、悪問が多く存在するテストは良質なテストとは言えないとしている。また、『良質さを論じる際に、テストについての理論的属性として、信頼性、妥当性を中心に、その他公平性や経済性などがあげられる。上記の悪問・奇問や良問についての議論はこれらの属性が多様に重なり合ってい

る。』とも続けており、テストの善し悪しは様々な要素があると述べている。永岡らは信頼性の高いことをもって良質なテストと判断し、実験結果において信頼性係数の値が高いテストを良質なテストであるとした。また、信頼性係数としてクロンバックの α 係数 [15] を用いている。

5.1.2 良い問題

良質なテストには良い問題（少なくとも悪問・奇問では無い問題）が多く含まれるということになる。そこで、本研究では、信頼性係数が高いテストに含まれる問題は良い問題である可能性が高いと定義し、誤答選択肢が自動生成された問題単体ではなく、誤答選択肢の自動生成を行った問題群を 1 つのテストとして扱い、信頼性係数の測定を行うこととした。

5.1.3 真の能力と信頼性

テストは受検者の能力を測定するためのものだが、受検するときの状況によって同じテストを同じ受検者が回答したとしても、点数に誤差が生じる。その人の本当の能力はテストでも完璧に測定することは出来ず、不明確なものだと言える [16]。テストを行うことで、能力はある程度測定出来るが誤差が生じるため、真の能力から離れた値になってしまう。これは仕方の無いことだが、なるべく誤差が少なくなり、個人の能力を真の能力に近い値で図れるテストが良いテストとされる。この誤差を信頼性と呼ぶ。信頼性の高い試験は真の能力からの誤差が小さく個人の能力を測定することができる。以下にこれらの計算式を示す。

$$\text{テストの得点} = \text{真の能力} + \text{誤差}$$

5.1.4 クロンバックの α

信頼性の値を表す指標として信頼性係数 α (クロンバックの α) という指標がよく利用される。 α の求め方は次の式になる。

$$\alpha = \left\{ \frac{\text{項目数}}{\text{項目数} - 1} \right\} \times \left\{ 1 - \frac{\text{各項目の分散の合計}}{\text{合計点の分散}} \right\}$$

α の値が 0.8 以上であれば信頼性が高い試験と言え、0.7 以上であればある程度信頼性は確保されていると言われて [17]。図 15 に信頼性係数の求め方を示す。

α の値を上げるためには以下の方法が挙げられる。

- ① 合計点の分散を大きくする (= 被験者の回答が、被験者同士バラバラになる)
- ② 各項目の分散の合計を小さくする (= ある項目に関して、被験者全員が同じような回答をする)
- ③ 項目数を増やす

このことから分かるように項目数を増やすことが一番簡単に α の値を上げる方法になるため、以降での試験の比較ではテストの項目数は統一することとした。項目数は 50 問あれば相応な α の値を得ることが出来る [18]

5.1.5 比較対象

本研究で対象としている「もりけん」で出題されたテストの信頼性を測定したところ、0.5~0.7 程度であった。本

研究では、作問委員会が誤答選択肢を作成したテスト（検定試験で使われたテスト）の信頼性係数と誤答選択肢を自動生成した問題を利用したテストの信頼性係数を比較し、同程度の信頼性であれば信頼性が確保されたと考える。

5.2 誤回答の収集対象

誤回答は、問題の難易度と受検者の能力差によってその内容が変化すると考えられる。そのため、誤回答から誤答選択肢を生成する場合、有効な誤答選択肢（正答と概念が近い単語）となる誤回答をどのような条件下で収集するかが課題となる。問題の難易度が低く、受検者の能力が高いと正答率は高くなり、誤回答の数が少なくなってしまう。逆に問題の難易度が高すぎる場合、回答が全く思い浮かばず、無回答数が増えてしまう。

試験難易度と受検者の能力の違いによる誤回答収集率の違いを測る実験を行った。結果を図 16 に示す。もりけん 1 級問題（1 問 1 答形式の問題）を 1 級の受検者が回答した結果が左側で、同様の問題を著者が所属する研究室の学生に回答させた結果が右側となる。もりけん 1 級は非常に難易度が高い試験で、2 級に合格していることが条件となっている 1 級受検者の正答率でさえ 31% となっている。それに対し、もりけんを受検したことの無い学生達の正答率は 1% にも満たない結果となった。学生にとっては難易度が高すぎた試験となり、誤回答よりも無回答率が増えているのが分かる。この内容から、問題の難易度と回答者の知識レベルに差がありすぎると誤回答を収集ができないことが分かった。

5.3 実験

もりけん 1 級問題（1 問 1 答形式の問題）に対する 1 級受検者の誤回答を活用して誤答選択肢の自動生成を行った。そして、四択形式の問題に変換した 1 級問題を著者が所属する学生に解答してもらった。解答結果からクロンバック α を求めたが、0.0007 となり、信頼性が極めて低い結果となった。これはもりけん 1 級の問題が学生にとって難易

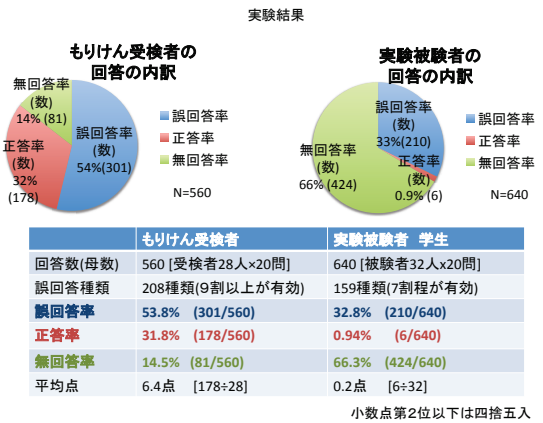


図 16. もりけん 1 級問題に対する二つの受検者グループの回答結果の差異

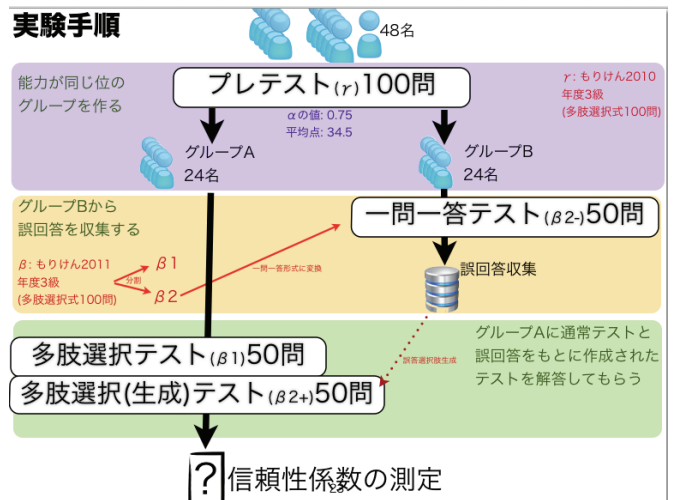


図 17. 実験手順

度の高すぎる問題であったためと考えられる。そこで、誤回答を収集する人と、誤答選択肢が自動生成された問題を解答する人の能力が同程度となるように実験を行った。もりけん 1 級問題に対して 1 級の受検者から誤回答収集を行い、1 級受検者と同程度の被験者を手配するのが望ましかったが、1 級受検者と同程度の能力を持った人はほとんどおらず、断念した。そこで岩手県立大学の学生 48 名を対象に実験を行った。図 17 に実験手順を示す。

5.3.1 プレテスト

被験者を誤回答収集群と自動生成後の問題解答群の 2 グループに分類するために、テスト γ を実施した。5.2 節で述べた実験から 1 級問題は学生には難しすぎると判断し、 γ は過去に出題されたもりけんの中で、難易度が低く（正答率が高く）、信頼性係数が高かったもりけん 2010 年度 3 級問題 100 問を採用した。実験前にプレテストを行い、平均点（34 点）が同じになるように A グループと B グループの二つに分割した。

5.3.2 誤回答の収集

誤回答を収集する問題にはプレテストとは別のテスト β

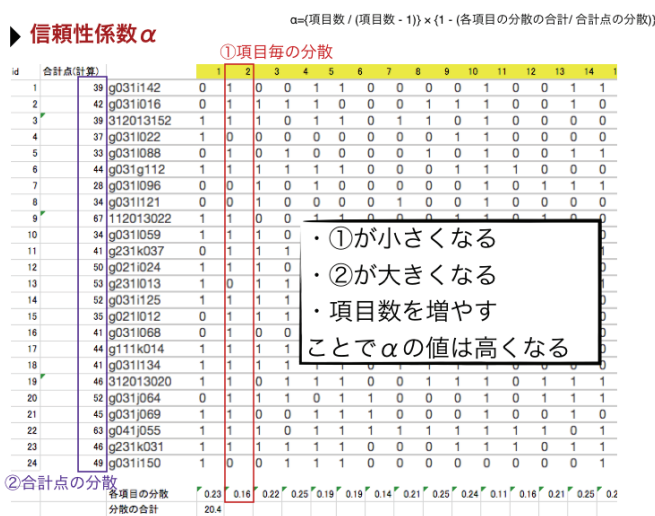


図 15. 信頼性係数の計算

を用意した。βはγと同じ条件になるように、γの次に正答率と信頼性係数が高かったもりけん 2011 年度 3 級 100 問とした。βを難易度が同じ位になるようにβ1とβ2の50問ずつに分割し、β2を一問一答形式問題へ変換した(β2-)。そして、この問題をグループ B に回答してもらい、誤回答を収集した。

5.3.3 誤回答から誤答選択肢生成を行ったテストの回答

グループ B から収集した誤回答を用いて誤答選択肢を生成し、β2-の問題を四択問題に変換したテストβ2+を作成した。そして、もともと1つの試験だったβ1とβ2+を1つのテストβ3としてグループ A に回答をしてもった。

5.3.4 結果

β3の回答をβ1とβ2+で分けて信頼性係数の分析を行った。もりけんの作問委員会が作成した問題のβ1と誤回答から誤答選択肢を生成して作成した問題β2+の信頼性係数が近い値になれば、誤回答から生成された誤答選択肢の信頼性が、作問委員会によって作成された誤答選択肢と同程度の信頼性があったと判断できる。その結果、β1の信頼性係数が0.6であったのに対してβ2+の信頼性係数は0.75となり、誤回答を用いて誤答選択肢を生成した問題の方が、信頼性が高い結果となった。以上より、誤答選択肢の生成に誤回答が活用できることが示唆された。

5.4 考察

5.3 節のような結果となった要因としては、誤回答を用いて誤答選択肢を生成した方が、少人数の作問委員会が誤答選択肢を作成するよりも、複数の観点で誤答選択肢を設定できるため、信頼性が上がったと考えられる。しかし一方では、表記の揺れによる紛らわしい選択肢が生成され、答えの選択にばらつきが生じたのではないかという見方もできる。

例えば、今回の実験では、誤回答を収集した1問1答形式問題の正答は単語の完全一致のみとしたため、表記のゆれによって誤回答となったものもあった。例えば、「盛岡市の姉妹都市の都市名を答えよ」という問題の場合、答えは「ビクトリア市」であったが、「ビクトリア」と答えた被験者の答えは誤回答と判断された。この場合、表記の揺れによって不正解(誤回答)になったことになる。このような回答での誤回答が多かったため、この問題に関しては「ビクトリア」が誤答選択肢として設定され、誤回答から誤答選択肢設定された問題にはビクトリア市とビクトリアが選択肢中に出てくる悪問となってしまった。紛らわしい選択肢設定により、グループ A の被験者は正解だと分かっていたにも関わらず不正解になった可能性も否定出来ない。

データ数は少ないが、今回の実験から、信頼性係数約0.7以上、正答率約35%程度の試験&受検者の組み合わせであれば有効な誤回答が収集できることが示唆された。

6. まとめと今後の課題

6.1 まとめ

本稿では試験作成現場の現状を紹介し、多肢選択形式問題の誤答選択肢の作成作業の負担軽減の重要性を述べた。その上で、1問1答形式の問題を多肢選択形式に変換することで作問負担を軽減する手法を提案し、誤回答、過去問題、JWOを用いた誤答選択肢自動生成手法を提案した。また、誤答選択肢の自動生成機能をもつ作問支援システム KenteiMakerを開発した。さらに、生成された誤答選択肢の評価実験を行い、誤回答を用いた誤答選択肢の生成の有効性を示唆した。

6.2 今後の課題

5.3.4 項のような紛らわしい選択肢に関しては、文字列類似度を考慮した手法により解決が見込めると考えている。正答と誤答選択肢の文字列類似度を測定し、近い場合は誤答選択肢の単語を入れ替えることで、紛らわしい選択肢をある程度解消出来る。例えば、ビクトリアとビクトリア市が選択肢に推薦された場合はビクトリアとビクトリア市で答えとなるビクトリア市を残し、ビクトリアは排除して別の単語を推薦するといった内容である。ただ、この例の場合は上手くいくが、例えばカナダの都市という誤答選択肢が推薦された場合、ビクトリア市とカナダの都市という単語は文字列類似度から見ると似ているとは言い切れない。しかし、これを選択肢として認めた場合、ビクトリア市もカナダの都市の一部であり、カナダの都市という単語も正解になるため、紛らわしい選択肢となる。一概に文字列類似度だけでは判断しきれないため、今後は意味も考慮した手法も検討したい。

また、KenteiMakerにおける誤答選択肢自動生成機能では、3章で述べた通り誤回答、類似問題、JWOの三つのリソースを用いて選択肢の自動設定を行っており、今回はその中でも誤回答の有効性の検証を行った。過去問題とJWOから生成した誤答選択肢の信頼性の検証も今後進めていきたい。さらに、本システムを使う事あらゆる試験作成現場の作業負担を軽減させることが出来る仕組みにまで発展させていきたい。

謝辞

この研究を進めるに当たり、盛岡商工会議所の皆さん、盛岡文化地層研究会の皆さん、佐々木研究室の教員の方々、プロジェクトメンバー、様々な方にお世話になった。携わった皆様に感謝致します。

7. 文献目録

- [1] 情報処理技術者試験. [Online]. <http://www.jitec.ipa.go.jp/>
- [2] TOEIC. TOEIC | コミュニケーション英語能力を測る世界共通のテスト. [Online]. <http://www.toeic.or.jp/>
- [3] 菅原遼介,高木正則,山田敬三 佐々木淳, "地域に帯する知識ベース充実へ向けた作問支援システムの設計,".
- [4] 菅原遼介, 高木正則, "誤回答を用いた誤答選択肢自動生成システムの開発と評価," 教育システム情報学会(JSiSE2012) 学生発表会 in 山形 JSiSE Conference for students in Yamagata, 2013.
- [5] 菅原遼介, 奥津翔太, 古館昌伸, 高木正則, 山田敬三, 佐々木淳, "多肢選択形式問題における誤答選択肢設定手法の検討," 情報処理学会 情報教育シンポジウム SSS2012, 2012.
- [6] 学びing 株式会社. Q ろいど 1-文章の穴掘り&四択自動生成. [Online]. <http://manabing.jp/service/qroid.html>
- [7] 学びing 株式会社.
<http://manabing.jp/services/autogenerate>.
- [8] Hiroshi Nakagawa Hoshino Ayako, *Assisting cloze test making with a web application.*: Proceedings of SITE 2007, 2007.
- [9] 岩手県立大学. もりけん web クイズ. [Online].
<http://sakumon.jp>
- [10] 香川 宏介, 森田 武史, 山口 高平 玉川 奨,
"日本語 Wikipedia オントロジーの Linked Open Data への取り組み,".
- [11] DBpedia Japanese. [Online]. <http://ja.dbpedia.org/>
- [12] LevenshteinV.I., "Binary Dodes Capable of Correcting Deletions, Insertions and Reversals," Doklady Akademii. Nauk SSSR, 163,4 1965.
- [13] 永岡慶三, "英語に関する「学力」と「テスト作成力」の関係の分析," 日本教育工学会,.
- [14] 福地 信義, "悪定義問題の解決~数理計画学~, "九州大学出版会, 2006.
- [15] Cronbach LJ, "Coefficient alpha and the internal structure of tests," Psychometrika 16 (3) 1951.
- [16] 若林俊輔, 無責任なテストが「落ちこぼれ」を作る.: 大修館書店, 1993.
- [17] J.M. Cortina, "What is coefficient alpha? An examination of theory and applications," 1993.
- [18] 妥当性と信頼性.
[Online].
<http://www.u-gakugei.ac.jp/~kishilab/validity-reliability.htm>