様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　６ 月 １２ 日現在

機関番号：３２６４４
研究種目：基盤研究(C)（一般）
研究期間：2020～2023
課題番号：２０Ｋ００８１４

研究課題名（和文）「Exam Corpus」の構築と、コーパスの言語テスト作成・改良への応用

研究課題名（英文）Construction of Exam Corpus and its application to making and improving language tests

研究代表者
宇佐美　裕子（Usami, Hiroko）

東海大学・教養学部・准教授

研究者番号：２０７３４８２５

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究では、1)2023年度までに出題された語彙・文法4択問題と会話問題を収集した「Exam Corpus」の構築と、2)コーパスの検索・分析技術を使用した英語検定テストの検証、
具体的には、(1)英語検定テストの語彙問題におけるCEFRレベルごとの頻出の語彙項目の抽出、(2)頻出の語彙問題(特に句動詞)に使用されている問題文を British National Corpusと比較して検証、(3)大学入試問題に出題された短い会話問題に使用されている英語を検証、(4)コーパスの検索・分析技術、特にsearch syntaxがどのように英語検定テストの語彙・文法問題の作成に応用できるか検証した。

研究成果の学術的意義や社会的意義
語彙・文法4択問題と会話問題を収集した「Exam Corpus」の構築と、コーパスの検索・分析技術を使用して英語検定テストを様々な点から検証した結果、
1)CEFRレベルごとにどの語彙項目が頻出か、2)語彙問題で使用されている問題文が実際に使用されている英語と異なること、3)短い会話問題に使用されている英語、出題されている項目、4)コーパスの検索技術を使用した効果的な語彙・文法問題の作成を検証・発見することができた。

研究成果の概要（英文）：In this research, I constructed the Exam Corpus containing multiple-choice vocabulary and grammar questions tested by 2023 and examined the English proficiency tests using corpus analytical techniques.
For example, (1) I extracted the vocabulary question items frequently tested across CEFR levels in the vocabulary questions in the English proficiency tests, (2) I examined the English used in the stems of vocabulary question items (e.g. phrasal verbs) frequently tested in the English proficiency tests and compared them to the British National Corpus, (3) I examined the English presented in short conversation questions in Japanese university entrance examination, (4) I examined how corpus analytical techniques (especially, search syntax) can be applied to make vocabulary and grammar questions in the English proficiency tests.

研究分野：コーパス言語学

キーワード：コーパス　言語テスト　日本の大学入試問題　語彙・文法問題　英検　TEAP　TOEFL　TOEIC

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様 式 C－19、F－19－1（共通）

1．研究開始当初の背景

①コーパスの英語教育への応用

コーパスは一般コーパスと学習者コーパスを含め、これまで語彙・文法の記述、辞書や語彙リスト、文法書の作成や改良に使用されており、教育の分野では、教材作成、コンピュータを用いた言語学習（CALL）、データ駆動の学習（data-driven learning）に応用されている。最近言語テストの分野にも応用されるようになったが、この分野の研究はまだ比較的新しく（投野, 2015; 金子, 2015）、研究する価値はあるものの、十分に発展していない。

②コーパスの言語テストへの応用

最初に言語テストの立場から、コーパスの言語テスト分野への応用に関するあらゆる可能性を提唱したのは Alderson (1996)であったが、コーパスを言語テストに応用できる理論的な提案のみであった。コーパスを言語テスト分野へ応用した実証研究は、当時 Cambridge ESOL の Barker（2006）であり、彼女が先駆者となって、コーパスの言語テストへの応用に関する研究は進められた。さらに、CEFR の枠組みで学習者コーパスを構築して、各 CEFR レベルの学習者の語彙・文法の特徴を研究し、英語教育、言語テストに応用する研究も盛んになった。実際、これまで、イギリスの Cambridge ESOL やアメリカの Educational Testing Service（ETS）はテストで使用されている英語の検証、問題作成、自動採点など様々な観点から積極的にコーパスを言語テスト研究に応用している。しかし、日本の大学入試問題でコーパスを用いた研究はまだ十分にされていない。日本の大学入試問題は、作問に不慣れな大学教員が他の業務で多忙な中、作成されているのが現状である。

③コーパスの言語テストへの応用の具体的研究

申請者は、大学入試問題の語彙・文法4択問題を収集した**「大学入試問題コーパス」（JUEEC）**を構築し、語彙・文法4択問題に使用されている英語、問われている語彙・文法項目について研究した。結果、語彙・文法4択問題には、一つの語彙・文法項目をテストするのに類似した英文が多用されていたり、英語圏で実際に使用されていない古い英文が頻出されているという結果が得られた（Usami, 2005; 2012; 2015）。さらに British National Corpus (BNC)を用いて問題文を改良し、Longman Learners' Corpus の日本人英語学習者の誤用を用いて錯乱肢を改良して学生に解かせた結果、誤用を用いて錯乱肢を改良した問題は非常に効果的であることが判明した。

2．研究の目的

日本の大学入試問題を検証・改良するため、日本人大学生英語学習者による英作文の誤用を収集した**「英作文コーパス」**を含め、他の英語検定テスト（英検、TEAP、Cambridge English Qualifications、Linguaskill、TOEFL ITP、TOEIC）の語彙・文法の過去問題を収集し、各問題にテスト名、技能（語彙/文法）、語彙・文法項目、CEFR レベル、難易度、弁別度、錯乱肢の選択率をダグ付した**「Exam Corpus」**を構築する。また、**コーパスの検索・分析技術**を用いて、各英語検定テストの語彙・文法問題で問われている語彙・文法項目、問題文と錯乱肢に使用されている英語を検証し、さらに問題文を変えた場合と、錯乱肢を学習者の誤用で改良した場合の各問題の難易度、弁別度、錯乱肢の選択率を既存の大学入試問題と比較し、**コーパスとコーパスの検索・分析技術が言語テスト分野（言語テスト作成・改良）にどのように応用できるか**を検証することを目的とする。

3．研究の方法

1）「Exam Corpus」の構築

①英語検定テスト（2000 年～2022 年の私立大学入試問題、英検の準 1 級～3 級、TEAP、Cambridge English Qualifications の A1～B2、Linguaskill、TOEFL ITP、TOEIC の Part 5）の語彙・文法の過去問題を収集し、1問1ファイルに入力する。

②各語彙・文法問題に、英語検定テスト名、技能（語彙/文法）、語彙・文法項目、CEFR レベル、難易度、弁別度、錯乱肢の選択率をタグ付する。（下記のサンプルファイルを参照）

③頻出の語彙・文法項目から誤用が多いものを選定し、英作文問題を作成して日本人大学生英語学習者の頻出語彙・文法項目の英作文の誤用を収集した「英作文コーパス」を構築する。

2）コーパスの検索・分析技術を使用した英語検定テストの検証

①「Exam Corpus」と、コーパスの検索・分析技術（ワードリスト、n-gram、キーワードリスト、コロケーション等）を使用して、各英語検定テストの語彙・文法問題における CEFR レベルごとの頻出の語彙・文法項目を検証する。

②同様に、問題文（stem）と錯乱肢に使用されている語（句）、難易度、弁別度、錯乱肢の選択率ごとに頻出の語彙・文法項目を検証する。

3）改良テストの作成と実施

①「Exam Corpus」とコーパスの検索・分析技術を使用して、テスト A（大学入試問題には頻出ではないが、他の英語検定テストに頻出の語彙・文法項目の問題）、テスト B（大学入試問題に頻出ではないが、他の英語検定テストに頻出の問題文、あるいは BNC2014 を用いて問題文を改良）、テスト C（「英作文コーパス」に収集された誤用を用いて錯乱肢を改良）をそれぞれ 10 問ずつ、合計 30 問の改良テストを作成する。

②統計ソフトウェア SPSS、Winsteps を使用して、既存の大学入試問題と改良テストの記述統計（mean, mode, median, range, standard deviation, skewness, kurtosis, histogram）、信頼性、相関、項目分析（難易度、弁別度、各錯乱肢の選択率）を量的に分析し、think aloud protocol の手法を用いて、改良テストの各問題を質的に分析する。

4．研究成果

1）Exam Corpus の構築

英語能力試験（2000 年～2023 年の私立大学入試問題、英検の準 1 級～3 級、TEAP、Cambridge English Qualifications の A1～B2、GTEC, TOEFL ITP、TOEIC の Part 5）の語彙・文法の過去問題と会話問題を収集して Exam Corpus を構築した。また各語彙・文法問題に、英語能力試験名、技能（語彙/文法）、語彙・文法項目、CEFR レベル、難易度、弁別度、錯乱肢の選択率をタグ付した。収集した問題数は以下の通りである。

技能 / テスト	大学入試問題	TEAP	英検	GTEC	TOEFL	TOEIC	CMB
会話問題	3,094	-	-	-	-	-	-
文法問題	10,876	0	405	0	176	903	0
語彙問題	9,442	60	1,995	92	0	463	380

2）コーパスの検索・分析技術を使用した英語能力試験の検証

英語能力試験で問われている語彙の CEFR レベル、問われている語彙項目、最も頻繁にテストされている語彙項目で提示されている問題文・解答・錯乱肢の観点から、英語能力試験の語彙問題に使用されている英語の妥当性の検証を行った。

具体的には、様々なコーパスを使用して、英語能力試験の語彙4択問題の、特に句動詞に使用されている英語を、1）頻繁にテストされている句動詞項目、2）頻繁にテストされている句動詞項目に提示されている問題文、の観点から調査した。結果、句動詞の中で put up with が最も頻繁にテストされているが、一般コーパスに見られる実際に使用されている英語では、それほど頻繁に使用されていないことが判明した。さらに、put up with の問題文に使用されている英語について、共通点はあったものの、

一般コーパスに見られる実際に使用されている英語と、英語能力試験で提示されている英語に幾つかの相違があったことも判明した。
さらに、Exam Corpus に収集された英語能力試験の短い会話問題において、どのような multiword unit が頻繁に、特徴的に n-gram とキーワードリストに現れているか、またどのような会話の状況がテストされているかを検証することを目的とした。結果、好み、思考、計画を問うもの、*how* と *mind* を使用した表現、申し出、依頼、道案内、座席の空き状況、電話での会話などが多く出題されていたことが判明した。
また、キーワードリストから、特に電話やお店での会話、*mind* を使用した許可を求める会話、道順を尋ねる会話が多く出題されていることも判明した。

3）コーパス検索・分析技術の言語テスト作成への応用
BNCweb や WordSmith Tools と#LancsBox などのコンコーダンサーで利用可能なコーパスの検索・分析技術（ワードリスト、キーワードリスト、コロケーション、クラスター等）がどのように言語テストに応用できるかを検証した。具体的には、頻繁にテストされている語彙問題や、文法問題でテストされている項目を特定し、文法問題で使用されている英語の真正について検証した。
また、BNCweb と#LancsBox で利用可能な search syntax を用いて、特定の接尾辞、接頭辞がついた語彙、様々な品詞の語彙、様々な活用の語彙、句動詞、受動態の問題をどのようにコーパス内で探すことができるかについて検証した。

５．主な発表論文等

〔雑誌論文〕　計6件（うち査読付論文　6件／うち国際共著　0件／うちオープンアクセス　6件）

１．著者名 Hiroko Usami	４．巻 53
２．論文標題 Examining targeted phrasal verb items and their stems in multiple choice vocabulary questions using authenticity and frequency in general corpora	５．発行年 2022年
３．雑誌名 東海大学教養学部紀要	６．最初と最後の頁 15-29
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Hiroko Usami	４．巻 1
２．論文標題 Applying the Exam Corpus to Inform Targeted Vocabulary in Synonym Questions in Japanese University Entrance Examinations	５．発行年 2022年
３．雑誌名 The Bulletin of Language Education Center Tokai University	６．最初と最後の頁 1-18
掲載論文のDOI（デジタルオブジェクト識別子） 10.18995/24367532.1.1	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Hiroko Usami	４．巻 3
２．論文標題 Applying Specialised Corpora to Language Testing: Development of the Exam Corpus (Vocabulary and Grammar Sections)	５．発行年 2021年
３．雑誌名 The Bulletin of International Education Center Tokai University	６．最初と最後の頁 1-13
掲載論文のDOI（デジタルオブジェクト識別子） 10.18995/24347337.3-1.1	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Hiroko Usami	４．巻 21
２．論文標題 Using Corpus Analytical Techniques for Language Testing: Identifying Frequent Question Items and Examining Authenticity	５．発行年 2023年
３．雑誌名 Hawaii International Conference on Education 2023 Conference Proceedings	６．最初と最後の頁 418-438
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Hiroko Usami	4．巻 54
2．論文標題 Examining short conversation questions' use of English with the Exam Corpus	5．発行年 2023年
3．雑誌名 東海大学教養学部紀要	6．最初と最後の頁 15-32
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Hiroko Usami	4．巻 22
2．論文標題 Using Corpus Search Syntax to Develop Vocabulary and Grammar Questions	5．発行年 2024年
3．雑誌名 Hawaii International Conference on Education 2024 Conference Proceedings	6．最初と最後の頁 341-361
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　3件）

1．発表者名 Hiroko Usami
2．発表標題 Using Corpus Analytical Techniques for Language Testing: Identifying Frequent Question Items and Examining Authenticity
3．学会等名 Hawaii International Conference On Education（国際学会）
4．発表年 2022年～2023年

1．発表者名 Hiroko Usami
2．発表標題 Examining the English used in multiple-choice vocabulary questions using authenticity presented in corpora
3．学会等名 Language Testing Forum (LTF)（国際学会）
4．発表年 2021年

１．発表者名 Hiroko Usami
２．発表標題 Using Corpus Search Syntax for Language Testing: Developing Vocabulary and Grammar Question Items
３．学会等名 Hawaii International Conference on Education（国際学会）
４．発表年 2023年～2024年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関