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　４ 年　 ９ 月 １３ 日現在

機関番号：１４３０２
研究種目：基盤研究(B)（一般）
研究期間：2018～2021
課題番号：１８Ｈ００６７６

研究課題名（和文）話題が語彙・文法・談話ストラテジーに与える影響の解明

研究課題名（英文）Clarifying the effects of topics on vocabulary, grammar, and discourse strategies

研究代表者
中俣　尚己（NAKAMATA, Naoki）

京都教育大学・教育学部・准教授

研究者番号：００５９８５１８

交付決定額（研究期間全体）：（直接経費）　　13,100,000円

研究成果の概要（和文）：本研究では話題が言語形式に与える影響を解明するため、２つの言語資源を作成した。一つは、既存の会話コーパスを話題別に分類し、そこから各話題に特徴的な語を抽出した『話題別日本語語彙表』である。もう一つは大学生のペアに５分ずつ、決められた15の話題について会話をしてもらい、それを録音・文字化した『日本語話題別会話コーパス：J-TOCC』である。
これらを分析したところ、おおよそ70%の語は何らかの話題に特徴的であると言えることがわかった。特に、文法を担う助詞・助動詞も半数以上が話題の影響を受けていることがわかった。談話標識も時に男女差や地域差以上に話題の影響を受けていることがわかった。

研究成果の学術的意義や社会的意義
日本語教育では内容・話題を重視する潮流が生まれており、本研究の成果物は話題ベースの教材開発、教育法開発に大いに役立つものと言える。すでに話題別の日本語単語帳が作成された。また、話題を統制された『日本語話題別会話コーパス』は、話題が言語形式に与える影響を観察することを目的に設計されたものであるが、話題統制されているからこそ、地域差や男女差を明らかにすることも可能である。何より、多様な話題についての2010年代の若者の話し言葉を収集、記録したという点でも価値を持つ。

研究成果の概要（英文）：In this study, two linguistic resources were created in order to clarify the effects of topic on linguistic form. One is the "Topic-Vocabulary table for Japanese Language Education." In order to build this table, we divided the existing conversation corpus by topic and extracted specific words of each topic. The second is "J-TOCC: Japanese Topic-oriented Conversation Corpus," in which pairs of university students were asked to engage in 5-minute conversations on 15 predetermined topics, which were then recorded and transcribed.
Analysis of the corpus revealed that 70% In particular, more than half of the function wprds that load grammar were also found to be effected by the topic. Discourse markers were also found to be effected by topic, sometimes more than gender or regional differences.

研究分野：日本語学

キーワード：話題　コーパス　言語資源　語彙表　話題精通度　特徴語　CEFR　文体

※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。

１．研究開始当初の背景

(1)　日本語教育の世界では、英語教育の影響などもあり、CBI(Content Based Instruction: 内容中心の教授法)や CLIL(Content--Language Integrated Learning: 内容と言語の統合的学習)といった内容志向型教授法(石川 2017)が取り入れられることが多くなってきた。

　また、初級日本語教科書も従来の文法積み上げ型シラバスから、トピックシラバスという話題で各課をまとめた内容のものも増えてきた。

(2)　このように，日本語教育では「内容」「トピック」に焦点が当てられることが増えてきているが，もちろん個々のトピックを通して「語彙」「文法」といった言語的な学習内容を学ぶことが肝要である。そのため，どのような話題にどのような語彙・文法が必要になるかといった情報を大規模データに基づいて調べることは教材開発において非常に重要なデータとなる．また，仮に文型積み上げ式の教材であっても，会話練習には話題は不可欠であるため，やはり非常に重要なデータとなる。

(3)　話題別の語彙表として山内（編）(2013)が存在していたが、文法を担う機能語は話題には依存しないとされていた。しかし、中俣(2016)や Nakamata(2019)は、接触場面会話コーパスを調査した結果、時間を表す「ている」「た」「昨日」「今」などが「ポップ・カルチャー」に特徴的であるなど、「文法も話題に依存する」と結論している。

２．研究の目的

(1)　母語話者の話題別コーパスを作成し、文法や談話ストラテジーも含めて、言語形式がどの程度話題の影響を受けているのかを明らかにする。この目的を達成するために２つの言語資源を構築する。

(2)　１つは、指定した話題について同じ時間だけの会話を収録した『日本語話題別会話コーパス：J-TOCC』である。話題を厳密に統制し、話題の影響を測定することができる。

(3)　もう１つは既存の自然会話コーパスである『名大会話コーパス』(藤村ほか 2011) を人手で話題ごとに分割し、その結果から特徴語を抽出した『話題別日本語語彙表』である。自然会話でよく出現する話題と、そこで使われる言語形式を知ることができる。

３．研究の方法

(1)　『J-TOCC』の作成はまず話題の選定から始まった。山内(2013)の 100 話題を出発点に、身近な話題を中心に、以下の４条件を考慮しながら 15 の話題を選定した。結果は表 1 の通りである。

表１　J-TOCC の話題（中俣ほか 2021a）

身の回りの話題	社会にもかかわる話題
01.食べること 02.ファッション 03.旅行 04.スポーツ 05.マンガ･ゲーム 06.家事 07.学校 08 スマートフォン 09.アルバイト 10.動物 11.天気	12.夢・将来設計 13.マナー 14. 住環境 15.日本の未来

　会話を録音する調査協力者は様々な話題について自然に話しやすい 20 代大学生の親しい友人同士の関係に限定した。関東地区と関西地区でバランスをとり，さらにそれぞれ「男女」「男男」「女女」の３グループを同じペア数だけ収録した．１グループあたり，20 ペア(40 名)が参加しており，話者の重なりはない．全体では 120 ペア，240 名となる。120 ペアにそれぞれ上記の 15 話題について５分ずつ話してもらったものを録音，文字化した。また，話者がそれぞれの話題についてどのぐらい詳しいかという「話題精通度」についても 5 段階で尋ね，付属データとして添付した。

(2)　『話題別日本語語彙表』を作成するため，まず既存の自然会話コーパスである『名大会話コーパス』を人手で話題ごとのサブコーパスに分割する作業を行った．話題については山内(編)(2013)を参考に 105 のタグセットを用意し，実際には 97 種類の話題に分割した．名大会話コーパスのどの行がどの話題に対応するかというデータについては中俣(2020)として，言語資源協会で配布している．その後，コーパス全体での頻度が 10 以上の 3,324 単語について，各サブコーパスにおける特徴度を LLR(対数尤度比)で示した Excel ファイルをウェブで公開した。いくつかの話題についての基礎統計量を表 2 に示す。

表 2　話題ごとの基礎統計量（中俣ほか 2020b を改変）

話題	延べ語数	異なり語数	特徴語数	特徴語例
食	95,250	5,565	236(最大)	美味しい，味，パン，入れる
パーティー	10,648	1,279	76	決める，日曜日，俺
家電	6,700	833	48(中央)	MD，マイク，テープ，CD
コンピュータ	6,294	867	29	ウイルス，使う，開く
税	260	119	1(最小)	こそ

４．研究成果

(1)　最終的な『J-TOCC』のサイズは 1 話題につき 10 時間, 15 話題合計で 150 時間となる．形態素解析の結果、およそどの話題でも 11 万語のデータを含んでいることがわかった。話題による差は見られない。一方、異なり語数にはもう少し差が見られた。また、話題精通度の平均値を見ると、大学生が最も得意な話題は「05 マンガ・ゲーム」であり，最も自信がなかった話題は「15. 日本の未来」である．しかし，身近な話題の中でも「02. ファッション」や「11. 天気」など，この値が低いものもあった．また，話題精通度と延べ語数（記号除）との相関係数は .132 とほとんど見られなかったが，異なり語数との相関係数は .600 と中程度であった。

(2)　『話題別日本語語彙表』は「話題から語を調べる」「語から話題を調べる」の 2 方向に活用可能である。解説動画を YouTube で公開した。例えば「食」に特徴的な語彙は表 3 のとおりである。様態の「そう」の他、可能表現も特徴的であった。

表 3 :「食」の話題に特徴的な語（抜粋・括弧内は LLR）（中俣ほか 2021c）

	名詞　(N=145)	動詞　(N=35)	形容詞　(N=26)	その他　(N=26)
1	味（308）	食べる（2,514）	美味しい（1,735）	此れ（82）
2	パン（216）	入れる（210）	好き（154）	そう-様態（77）
3	コーヒー（142）	飲む（140）	固い（88）	まー（71）
4	鍋（140）	頂く（135）	甘い（88）	どうぞ（70）
5	御飯（124）	煮る（134）	御腹一杯（71）	はい（69）

一方、語がどのような話題に特徴的かという観点からは、全体の 3/4 の語が何らかの話題で特徴語となっていることがわかった。これは品詞によって割合が異なり、図示すると図 1 のようになる。副詞・接続詞はあまり話題に依存しないこともわかった。

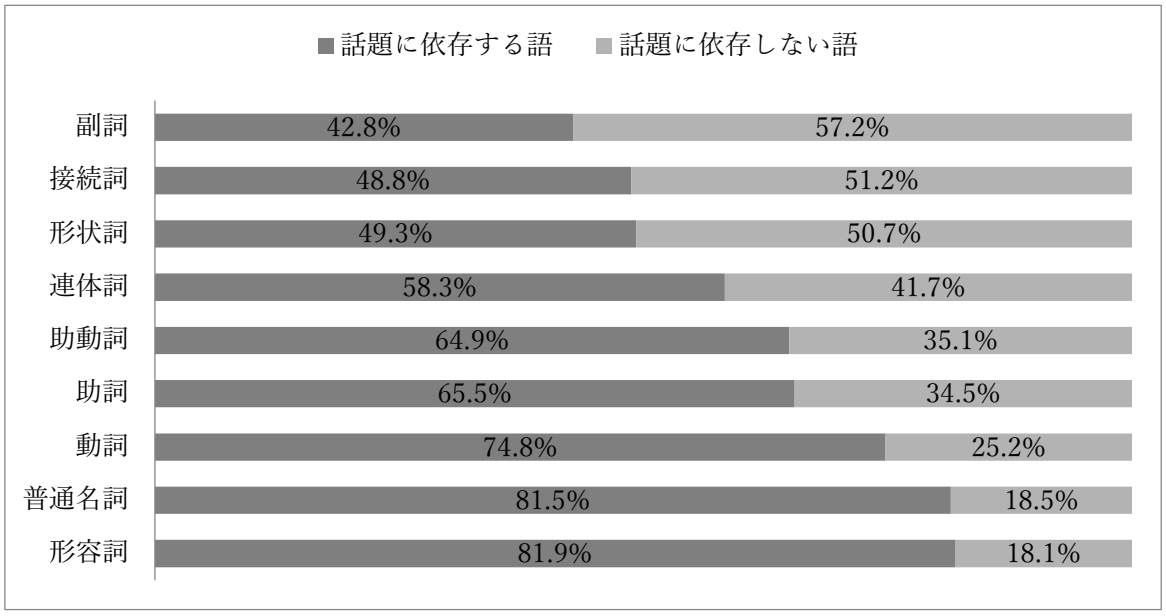

図 1：品詞ごとの話題に依存する語の割合（中俣ほか 2021c）

語と話題の関係が明らかになると、例えば類義語の説明に役立つと考えられる。表 4 は類義形容詞の比較であるが、「ハンサム」「美しい」はそれが使われる話題が非常に限られていることがわかる。他の話題で用いると、違和感をもたれるかもしれない。

表 5：類義形容詞の特徴話題の比較（中俣ほか 2021c）

語	特徴話題
ハンサム	映画・演劇
格好良い	芸能界，恋愛，写真，衣
綺麗	芸能界，写真，工芸，植物，旅行，建設・土木，町，自然・地勢，死
美しい	マナー・習慣
可愛い	芸能界，写真，衣，美容，工芸，ものづくり，動物

(3)　研究背景としてあった「文法も話題に依存するのか？」という問いの検証のためにアスペクト・マーカーの「てる」を取り上げる。日中『Skype 会話コーパス』ではこの形式は「ポップ・カルチャー」の特徴語であったが、『J-TOCC』においては最もよく出現するのは「05.マンガ・ゲーム」であり次に「08.スマートフォン」が多い。さらに、『話題別語彙表』では「映画・演劇」「音楽」「ヒト」「友達」に多く、３つの独立した言語資源で一貫した結果が得られた。なお、縮約形「てる」と非縮約形「ている」の比率は『J-TOCC』では 98:2 にも達しており、また「ている」はやや硬い話題に多くなることがわかった。

(4)　2022 年 3 月 20 日にシンポジウム「話題とコーパスと日本語教育」を開催、成果物の解説を行うとともに、成果を利用した研究成果の発表を行った。その成果の一部を列挙する。なお、この成果は論文集として 2023 年に公刊予定である。

- 「～かな」のような自問発話は「01.食」よりも「15.日本の未来」に多い。東日本でよく使われ、形式にも東西差が見られる。
- 「まあ」も「15.日本の未来」に多い。また、男性がよく使用している。
- 間投助詞「さ」も「03.旅行」より「15.日本の未来」に多い。話題差は地域差・性差よりも大きく影響する。終助詞の「さ」はほとんど見られない。
- オノマトペは身近な話題のほうがよく使われる。また、「西日本」「女性」の頻度が高い。
- 「ことがある」という助詞ありパターンと「ことある」という助詞なしパターンを比較すると、「01 食べること」「05 マンガ・ゲーム」「09 アルバイト」は助詞なしパターンが多く，「14 住環境」「15 日本の未来」は助詞ありパターンが多い
- 提題形式について、「は」は「06 家事」「12 夢・将来設計」「14 住環境」で多用され、「も」は「14 住環境」「15 日本の未来」で多用されていた。
- 話題によって、経験を語る「た」と話題精通度の相関に差が見られた。
- 社会にかかわる話題は統語的複雑さが高く、身近な話題は統語的複雑さが低い。
- 成果物を用いて、授業におけるタスクで使用する機能語・実質語を絞り込んだり、真正性のある会話例を提示したり、OPI (Oral Proficiency Interview)と呼ばれる口頭能力テスト
  において話題ごとにレベルチェックのためのタスクを作成することができる。

(5)　『J-TOCC』は話題の影響を見るために設計されたコーパスであるが、話題の統制を利用して性差や地域差をこれまでのコーパス以上に厳密に測定することにも利用できる。また、2010 年代の大学生の会話を様々な話題について記録したという点においても価値がある。コーパスは 2021 年３月に公開されたが、人工知能方面での利用申請は多い。

(6)　教育への直接的応用として、２つの成果物を利用して単語を話題別に半自動で分割して作成した単語帳である中俣(2021b)を出版した。

参考文献

藤村逸子・大曽美恵子・大島ディヴィッド義和(2011). 会話コーパスの構築によるコミュニケーション研究.藤村逸子・滝沢直宏(編). 言語研究の技法：データの収集と分析, ひつじ書房, pp.43-72.

石川慎一郎(2017). ベーシック応用言語学. ひつじ書房.

中俣尚己(2016). 学習者と母語話者の使用語彙の違い―『日中 Skype 会話コーパス』を用いて―.日本語／日本語教育研究, 7, pp.21-34.

Nakamata, Naoki.(2019). Vocabulary Depends on Topic, and So Does Grammar. Journal of Japanese Linguistics, 35(2), pp.213-234.

中俣尚己(2020). GSK2020-B 自然会話コーパス話題アノテーション情報. https://www.gsk.or.jp/catalog/gsk2020-b

中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣(2021a). 『日本語話題別会話コーパス：J-TOCC』. 計量国語学, 33(1), pp.205-213.

中俣尚己・加藤恵梨・小口悠紀子・小西円・建石始(2021b). ミニストーリーで覚える日本語能力試験ベスト単語 2100. ジャパンタイムズ出版.

中俣尚己・小口悠紀子・小西円・建石始・堀内仁(2021c). 自然会話コーパスを基にした『話題別日本語語彙表』. 計量国語学, 33(3), pp.194-204.

山内博之(編)(2013). 実践日本語教育スタンダード. ひつじ書房

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 中俣尚己・小口悠紀子・小西円・建石始・堀内仁	４．巻 33-3
２．論文標題 自然会話コーパスを基にした『話題別日本語語彙表』	５．発行年 2021年
３．雑誌名 計量国語学	６．最初と最後の頁 pp.194-204
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣	４．巻 33-1
２．論文標題 日本語話題別会話コーパス：J-TOCC	５．発行年 2022年
３．雑誌名 計量国語学	６．最初と最後の頁 11-21
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 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 橋本直幸	４．巻 86
２．論文標題 「日本語教科書読み物データベース」の作成と公開	５．発行年 2022年
３．雑誌名 文藝と思想	６．最初と最後の頁 45-55
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 無
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 Nakamata, Naoki	４．巻 35-2
２．論文標題 Vocabulary Depends on Topic, and So Does Grammar	５．発行年 2019年
３．雑誌名 Journal of Japanese Linguistics	６．最初と最後の頁 213-234
掲載論文のDOI（デジタルオブジェクト識別子） 10.1515/jjl-2019-2011	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計22件（うち招待講演　16件／うち国際学会　3件）

１．発表者名 中俣尚己
２．発表標題 間投助詞「さ」と話題の関係
３．学会等名 第46回社会言語科学会研究大会（招待講演）
４．発表年 2021年

１．発表者名 加藤恵梨
２．発表標題 大学生が日常会話で使用する自称詞と対称詞について
３．学会等名 The 28th　Princeton Japanese Pedagogy Forum（国際学会）
４．発表年 2022年

１．発表者名 中俣尚己
２．発表標題 話題が語彙・文法・談話ストラテジーに与える影響を解明するための ２つの成果物
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 小西円
２．発表標題 話題・地域の異なりから みた自問発話
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 森篤嗣
２．発表標題 話題精通度と言語表現の出現傾向の 関係：タ形を例として
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 堀内仁
２．発表標題 話題と統語的複雑さ
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 加藤恵梨
２．発表標題 大学生の自然会話にみられる「まあ」―話題による使用の特徴に着目してー
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 太田陽子
２．発表標題 大学生の雑談におけるオノマトペの 使用傾向ー話題・地域・男女
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 橋本直幸
２．発表標題 日本語教材における「話題」の偏り
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 清水由貴子
２．発表標題 話題と無助詞現象の関わりについて
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 建石始
２．発表標題 類義語分析に「話題」は使える
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 澤田浩子
２．発表標題 話題と話題提示形式の関係
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 小口悠紀子
２．発表標題 J-TOCCと話題別日本語語彙表を中級レベルの日本語授業にどう活かすか
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 石川慎一郎
２．発表標題 コーパス研究における ジャンル・トピック・シチュエーション・タスク
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 山内博之
２．発表標題 話題を制する者は日本語教育を制す
３．学会等名 科研費成果物公開記念シンポジウム「話題とコーパスと日本語教育」（招待講演）
４．発表年 2022年

１．発表者名 中俣尚己
２．発表標題 自然会話コーパスを元にした話題別語彙表の作成
３．学会等名 第11回　日本語実用言語学国際会議(ICPLJ 11)（国際学会）
４．発表年 2020年

１．発表者名 中俣尚己・太田陽子・加藤恵梨・澤田浩子・清水由貴子・森篤嗣
２．発表標題 「日本語話題別会話コーパス:J-TOCC」の構築
３．学会等名 第11回　日本語実用言語学国際会議(ICPLJ 11)（国際学会）
４．発表年 2020年

１．発表者名 中俣尚己
２．発表標題 話し言葉における助詞の出現頻度に対する話題の影響
３．学会等名 日本語学会2021年度春季大会
４．発表年 2021年

１．発表者名 森篤嗣
２．発表標題 話題に対する知悉度と言語表現の出現傾向の関係
３．学会等名 日本語学会2021年度春季大会（招待講演）
４．発表年 2021年

１．発表者名 中俣尚己・建石始・堀内仁・小西円・山本和英
２．発表標題 自然談話コーパスに対する話題アノテーションの試み
３．学会等名 言語処理学会第26回年次大会
４．発表年 2020年

１．発表者名 中俣尚己・建石始・堀内仁・小西円
２．発表標題 『名大会話コーパス』で話されている話題の計量的分析―話題バイグラムを用いて―
３．学会等名 日本語教育学会2020年度春季大会
４．発表年 2020年

１．発表者名 中俣尚己
２．発表標題 学習者コーパスを使った研究の方法―誤用分析を超えて―
３．学会等名 中国語話者のための日本語教育研究会（招待講演）
４．発表年 2019年

〔図書〕　計1件

１．著者名 中俣 尚己（著，編集），加藤 恵梨（著），小口 悠紀子（著），小西 円（著），建石 始（著），	４．発行年 2021年
２．出版社 ジャパンタイムズ出版	５．総ページ数 292
３．書名 ミニストーリーで覚える JLPT日本語能力試験ベスト単語N3 合格2100	

〔産業財産権〕

〔その他〕

データベース	中俣尚己Webサイト http://nakamata.info/database/

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
研究分担者	山内　博之 (Yamauchi Hiroyuki) (20252942)	実践女子大学・文学部・教授 (32618)	
研究分担者	太田　陽子 (Ota Yoko) (20373037)	一橋大学・森有礼高等教育国際流動化機構・教授 (12613)	
研究分担者	森　篤嗣 (Mori Atsuhsi) (30407209)	京都外国語大学・外国語学部・教授 (34302)	
研究分担者	橋本　直幸 (Hashimoto Naoyuki) (30438113)	福岡女子大学・国際文理学部・准教授 (27103)	
研究分担者	堀内　仁 (Horiuchi Hitoshi) (40566634)	国際教養大学・国際教養学部・准教授 (21402)	
研究分担者	小西　円 (Konishi Madoka) (60460052)	東京学芸大学・留学生センター・准教授 (12604)	
研究分担者	清水　由貴子 (Shimizu Yukiko) (60735851)	聖心女子大学・現代教養学部・講師 (32631)	
研究分担者	澤田　浩子 (Sawada Hiroko) (70379022)	筑波大学・人文社会系・准教授 (12102)	

６．研究組織（つづき）

	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
研究分担者	建石　始 (Tateishi Hajime) (70469568)	神戸女学院大学・文学部・教授 (34510)	
研究分担者	小口　悠紀子 (Koguchi Yukiko) (70758268)	広島大学・人間社会科学研究科（教）・准教授 (15401)	
研究分担者	加藤　恵梨 (Kato Eri) (70770311)	大手前大学・現代社会学部・准教授 (34503)	
研究分担者	山本　和英 (Yamamoto Kazuhide) (40359708)	長岡技術科学大学・工学研究科・准教授 (13102)	

	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
研究協力者	石川　慎一郎 (Ishikawa Shin'ichiro)		
研究協力者	茂木　俊伸 (Mogi Toshinobu)		

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