

令和 5 年 5 月 31 日現在

機関番号：32689

研究種目：基盤研究(B)（一般）

研究期間：2019～2022

課題番号：19H01273

研究課題名（和文）論理的文章作成を支援するウェブ型自動処理システムの開発と普及に関する研究

研究課題名（英文）Research on the Development and Dissemination of a Web-based Automatic Processing System for Supporting Logical Composition of Texts

研究代表者

李 在鎬（LEE, JAEHO）

早稲田大学・国際学院（日本語教育研究科）・教授

研究者番号：20450695

交付決定額（研究期間全体）：（直接経費） 7,950,000円

研究成果の概要（和文）：本研究は計量言語学の方法を用いて、文章の論理性を自動評価するシステムを開発し、公開することを目的とする。研究の初年度には、文章の論理性のコア要素として接続表現に関する調査分析を行った。具体的にはこれまでの日本語学の研究成果を取り入れ、接続表現のリストを作成した。75見出しの接続表現を自動抽出する機能を <https://jreadability.net/> で実装し、日本語学習者、日本語教育関係者に対するウェブサービスとして提供した。そして、文章の論理性を自動判定する判別式を実装しました。入力文章に対して、コーパスから学習し、線形回帰分析によって導きだされた判別式を実装した。

研究成果の学術的意義や社会的意義

これまで、文章の論理性の判断は、熟練した人間の評価者のみができることと考えられてきたが、本研究では、大規模なテキストデータを統計的に処理することで、論理性を計算するモデルを開発した点で学術的意義は大きい。そして、その計算モデルをウェブシステム「<https://jreadability.net/jwriter/>」において実装し、誰でも利用できる環境を構築した。それにより、ビジネスパーソンを始めとする様々な人々が利用できるため、その社会的意義や波及効果は大きく、人工知能時代の教育や学習を後押しするものになった。

研究成果の概要（英文）：The aim of this research is to develop and publish a system that automatically evaluates the logicity of sentences using the methods of quantitative linguistics. In the first year of the research, we conducted an investigation and analysis of conjunction expressions, which are core elements of the logicity of sentences. Specifically, we created a list of conjunction expressions, incorporating the results of previous Japanese language studies. We implemented a function on <https://jreadability.net/> that automatically extracts 75 heading conjunction expressions, which is provided as a web service for learners and teachers of Japanese. We also implemented a discriminant equation to automatically determine the logicity of sentences. We implemented this equation, which was derived from linear regression analysis after learning from a corpus, on the input sentences.

研究分野：計量言語学

キーワード：計量言語学 日本語教育 作文教育 自然言語処理技術の教育応用 自動採点 文章の計算モデル

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

留学生30万人計画の旗揚げにともない、様々な教育機関において独自の留学生教育がなされている。とりわけ大学などの高等教育機関での修学において「書く力」は不可欠な要素であるが、これに関わる日本語教育機関の取り組みは必ずしも十分ではない。その理由として文章力というのは、読む人（評価する人）によって評価基準が異なるため、「良い文章」に対する唯一無二の正解を決めるのは非常に難しい。となると、万人にとって「良い文章」という正解を設け、それをゴールにしたシラバス化することも難しく、「書く力」をつけるための一貫性のある教育プログラム化も難しい。

こうした現状に対して、本研究グループでは2017年頃から計算モデルによる自動処理の可能性を示した。とりわけ重回帰分析により、学習者作文の能力値を判定する手法を提案している。これらの研究は、コンピュータ処理であるため、評価の一貫性が保たれることから信頼性の観点からは注目すべきものである一方で、解析精度が十分ではないことや評価の観点が教師の教育活動と一致していない点が問題とされている。こうした課題を受け、本研究では、次の2点を目指すことで現状の改善をはかる。1) テキストデータの統計的分析手法を用いて「良い文章」の包括的評価スケールとして文章の論理性を評価する基準を作成すること、2) 文章の論理性を自動判定するウェブシステムを構築し、世界中の日本語学習者に対する教育支援を行うこと。

## 2. 研究の目的

本研究の目的は以下の2点である。

- 1) 文章の論理性を自動評価するための計算モデルを開発すること。
- 2) 計算モデルをウェブシステムに実装し、日本語学習者の自律学習に役立てること。

以上の研究目的が達成されると、論理的作文の自動評価システムができる。このシステムは論理性の度合いを判断するとともに、論理性が欠如している原因を見つけ、自動生成したコメントを学習者にフィードバックする。こうしたシステム化をとおして、日本語教育における運用力評価の新たな潮流を作り出すことを最終ゴールにする。

こうした作業の学術的な意義として、次のことが考えられる。これまで日本語学および日本語教育においても「良い文章」や「論理的文章」のモデルは示されてきたが、それを評価するのは基本的には教師の主観であり、客観的な指標で一貫性・信頼性のある評価はなされてこなかった。また、文章の論理性に関する議論においても部分的な側面に注目するものが多く、全体としての姿を明らかにしたものは少ない。これまでの先行研究に対して、本研究は、以下のアプローチをとっている点で、独創的と言えるし、関連分野への波及効果が期待される。

- 1) 大規模な文章の論理性に関するアノテーションデータで教師の主観を定量的に捉え、コンピュータモデルでもって教師の主観を定式化する。→自然言語処理・計算科学へ貢献
- 2) 論理性の決め手となる言語的要素のあらゆる側面を集め、それぞれの要素に統計的な重みを与え、論理性を接続詞や文末モダリティや文長といった言語要素の分散表現として捉える。→計量言語学へ貢献
- 3) 学習者の作文に見られる論理性の欠如に対して、コンピュータが自動でフィードバックを行い、より良い文章執筆のためのアドバイスを行う。→日本語学習者へ教育支援

### 3. 研究の方法

本研究では、次の3つのステージで研究を行う。

■**ステージ 1:【論理性の計算モデル構築】** 研究代表者および分担者がこれまで構築してきた作文データベースをもとに、計算モデル作成のための学習データを構築する。大規模な作文のデータベースを構築し、それらに対して、複数の教師が論理性の度合いをアノテーションする。統計分析では、作文のテキスト特徴量を独立変数に、複数の教師が行った論理性判定度を従属変数にし、テキスト特徴量から論理性判定を予測する統計分析（判別分析、重回帰分析、決定木分析）を行う。

■**ステージ 2:【ウェブシステム開発】** ステージ 1 で作った計算モデルをウェブシステムに実装する。ウェブシステムへの実装は、研究代表者および分担者が昨年までの基盤研究によって開発した jWriter を改良する形で行う(図 1)。

jWriter では、外部からテキストを受け取り、その作文が初級から超級のレベルに相当するかを判定する。本システムは 2017 年に公開して以来、すでに 15000 件以上のアクセスがある。一方、解析精度が 70%にしか満たないことと文章の内容面に対する評価は全く行っていないという課題がある。これを受け、本研究では、計算モデルや出力方法を大幅に改良する予定である。とりわけ、ステージ 1 の計算式による論理的文章の度合いを評価する仕組みと言語的要素（語彙の多様性、漢語の使用率、長文作成力）を評価する。これに加え、文章を改善するためのコメントをフィードバックする。

■**ステージ 3:【広報や啓発】** 本研究は、ウェブシステム開発だけが目的ではなく、作ったシステムが日本語教育の現場に広がり、日本語教育分野における作文教育が充実していく様子が確認できることが、プロジェクトの最終ゴールであると考えている。そのため、研究の開始時からワークショップや学会広報をとおして、研究の全期間において広報や啓発活動を積極的に展開する。日本語教師や学習者は学会には来ない。広報の方法を学会だけに限定しない。具体的にはアジア、アメリカ、欧州の各地でワークショップを行う。また、専門家に向けては日本語教育学会をはじめ、言語処理学会、計量国語学会をとおして、広報および論文執筆、書籍刊行を行う。以上をとおして、様々な立場の日本語教師にプロジェクトの成果物を紹介し、共同利用を促す。

### 4. 研究成果

ステージ 1 の研究成果として、文章の論理性を自動評価する計算モデルを開発した。計算モデルの作成は、大規模なテキストデータを学習する形で行った。具体的な計算モデルは、次のとおりである。「 $y=3.099+(\text{普通名詞率} \times -2.837)+(\text{初級前半語率} \times 4.292)+(\text{接続詞率} \times -49.773)+(\text{動詞率} \times 4.281)+(\text{外来語率} \times -6.17)+(\text{代名詞率} \times -10.525)+(\text{助動詞率} \times 0.279)+(\text{形状詞率} \times -0.691)$ 」。

この計算モデルを使えば、入力テキストの論理性は、4段階「優レベル、高レベル、中レベル、低レベル」で評価される。

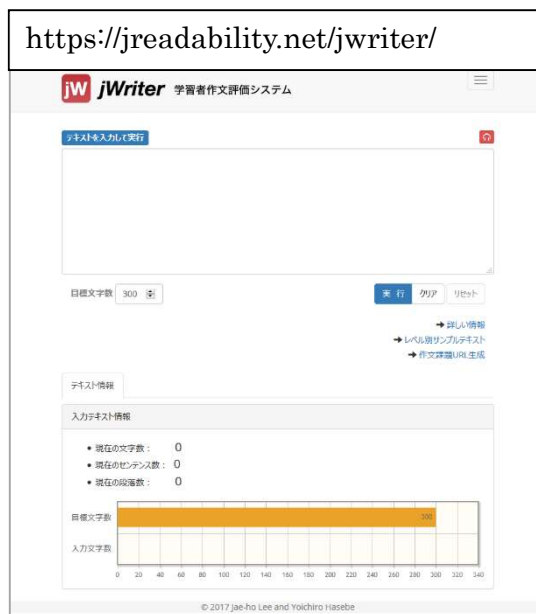


図 1 jWriter の入力画面

ステージ2の研究成果として、ステージ1で作成した計算モデルをウェブに実装した。具体的には1000文字以上のテキストに対して、論理性を評価するシステムが完成した。以下の図2は、本報告書の1ページ目の文章を入力した場合の結果である。



図2 論理性の判定結果 (<https://jreadability.net/jwriter/>)

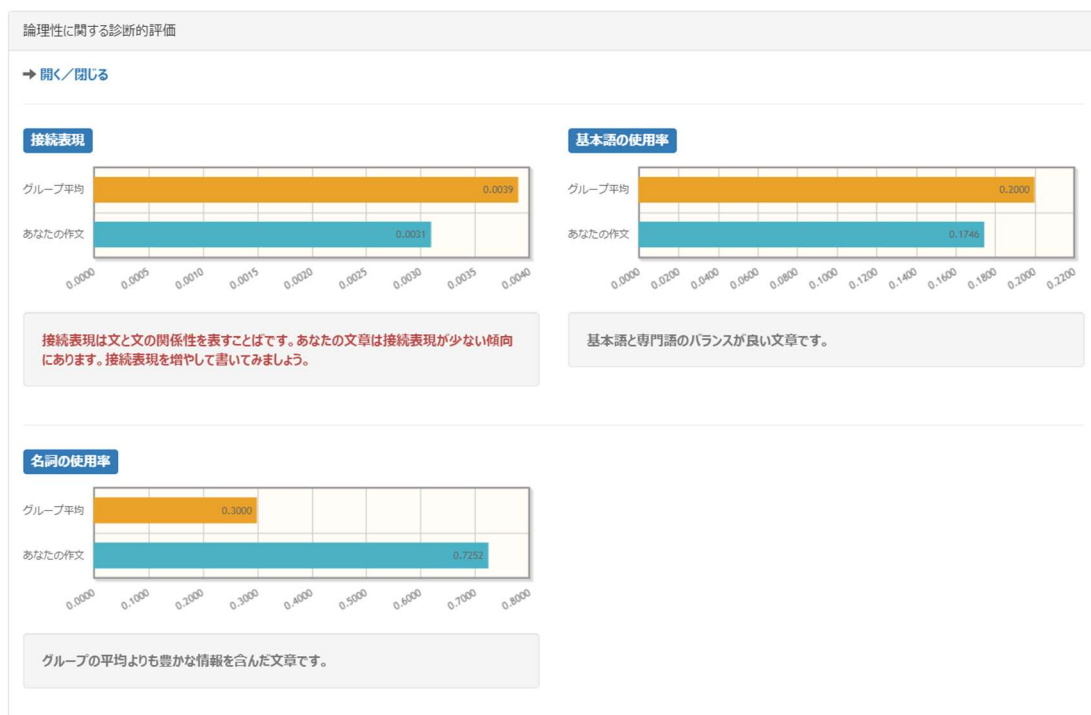


図3 論理性に関する判定結果の詳細および診断的評価の結果

論理性の判定は、図2の4段階の包括的な評価（優～低のいずれか）のほかに、図3のような診断的な評価を行っている。図3の診断的評価では、グループの平均をグラフで可視化するほかに、フィードバックコメントも表示するようにしており、文章の推敲に役立てる仕様になっている。こうしたシステムを教育に活用することで、自律的な文章の執筆や推敲ができると考えている。

ステージ2のウェブシステムへの実装は、論



図4 接続評価の可視化機能

理性のほかに、論理性の判断に重要な役割を果たす接続表現の使用状況を可視化する機能も実装した。テキストを入力すると、図4のように接続表現をカテゴリ別にハイライトする機能を追加した。

ステージ3の広報や啓発の活動として、日本語教育の関連学会のほかに、言語処理学会、教育工学関連の学会、言語心理関連の学会においてシステムを紹介した。学会発表の成果は、<https://researchmap.jp/jhlee/> を参照してほしい。学会発表のほかに、アフリカやヨーロッパの日本語教師会において講演会、ワークショップを行った。これをとおして、利用者のフィードバックを収集し、システムの更新を行った。具体的には、エジプト・カイロ (2022.12)、ベルギー・ルーヴェン (2023.1)、ハンガリー・ブダペスト (2023.2)、ポーランド・ポズナン (2023.5)、スペイン・クエンカ (2023.6)、ポルトガル・ポルト (2023.6) において、本研究が開発した自動採点システムを紹介する目的でセミナーを行った。セミナーを通して、日本語教師や日本語学習者のフィードバックを得ることができ、システムの部分的な修正を行ったし、広報活動も兼ねていたため、共同利用を活性化できた。

## 5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 4件 / うち国際共著 1件 / うちオープンアクセス 1件）

1. 著者名 李在鎬	4. 巻 32
2. 論文標題 ICT を利用した読解授業について	5. 発行年 2020年
3. 雑誌名 日本語教育連絡会議論文集	6. 最初と最後の頁 15-22
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 李在鎬	4. 巻 32-5
2. 論文標題 多人数調査法の現在（8）：Googleフォームによるオンライン調査	5. 発行年 2020年
3. 雑誌名 計量国語学	6. 最初と最後の頁 296-307
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 李在鎬, 伊東 祐郎, 鎌田 修, 嶋田 和子, 坂本 正, 六川 雅彦, 由井 紀久子	4. 巻 23
2. 論文標題 口頭能力と自己評価の関連性について	5. 発行年 2020年
3. 雑誌名 2019日本語教育シンポジウム第23回AJEヨーロッパ日本語教育シンポジウム報告・発表論文集	6. 最初と最後の頁 310-317
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 李在鎬	4. 巻 32-7
2. 論文標題 日本語教育学の課題に対して計量分析は何ができるか	5. 発行年 2020年
3. 雑誌名 計量国語学	6. 最初と最後の頁 372-386
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 伊集院 郁子, 李 在鎬, 小森 和子, 野口 裕之	4. 巻 23
2. 論文標題 評価コメントに見られる意見文評価の様相 共起ネットワーク及びコレスポネンス分析に基づく考察	5. 発行年 2020年
3. 雑誌名 第二言語としての日本語の習得研究	6. 最初と最後の頁 26-43
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 Lee, Jae-Ho, Hasebe, Yoichiro	4. 巻 5
2. 論文標題 Quantitative Analysis of JFL Learners' Writing Abilities and the Development of a Computational System to Estimate Writing Proficiency	5. 発行年 2020年
3. 雑誌名 Learner Corpus Studies in Asia and the World	6. 最初と最後の頁 105-120
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 LEE, Jae-ho, HASEBE, Yoichiro	4. 巻 1
2. 論文標題 Readability measurement of Japanese texts based on levelled corpora.	5. 発行年 2020年
3. 雑誌名 The Japanese Language from an Empirical Perspective	6. 最初と最後の頁 143-168
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 該当する

1. 著者名 李 在鎬	4. 巻 1
2. 論文標題 学習者の語彙使用は習熟度を反映しているか：学習者コーパスの定量的分析	5. 発行年 2019年
3. 雑誌名 学習者コーパスと日本語教育研究 (くろしお出版)	6. 最初と最後の頁 87-104
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 李在鎬	4. 巻 32-3
2. 論文標題 BCCWJ の学校教科書コーパスの計量的分析 日本語教育のためのリーダビリティと語彙レベルの分布を中心に	5. 発行年 2019年
3. 雑誌名 計量国語学	6. 最初と最後の頁 147-162
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 李在鎬	4. 巻 1
2. 論文標題 学習者作文と自他の使用頻度	5. 発行年 2020年
3. 雑誌名 自動詞と他動詞の教え方を考える (くろしお出版)	6. 最初と最後の頁 43-56
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 -

〔学会発表〕 計12件 (うち招待講演 2件 / うち国際学会 1件)

1. 発表者名 伊集院郁子・李在鎬・青木優子・長谷部陽一郎・村田裕美子
2. 発表標題 複数のコーパス分析に基づく接続詞使用と作文トピックの関係性
3. 学会等名 第22回専門日本語教育学会研究討論会
4. 発表年 2020年

1. 発表者名 三谷彩, 滝島雅子, 李在鎬
2. 発表標題 学術論文の構成要素に関する計量的分析 日本語教育学分野の研究論文を対象に
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 李在鎬, 伊集院郁子, 青木優子, 長谷部陽一郎, 村田裕美子
2. 発表標題 I-JASを用いた習熟度と接続詞の使用に関する調査：論理的文章執筆の支援システムの構築に向けて
3. 学会等名 計量国語学会第64回大会
4. 発表年 2020年

1. 発表者名 村田 裕美子, トリチコヴィッチ・ディブナ, 李 在鎬
2. 発表標題 異文化間能力の育成を目指す計量テキスト分析：ドイツ・セルビア・日本の学生を対象に
3. 学会等名 計量国語学会第64回大会
4. 発表年 2020年

1. 発表者名 李在鎬
2. 発表標題 オンライン日本語教育の過去, 現在, 未来：ICT肯定論からデータ科学へ
3. 学会等名 韓国日語日文学会2020年度秋季大会（招待講演）
4. 発表年 2020年

1. 発表者名 李在鎬
2. 発表標題 機械学習に基づく話し言葉と書き言葉の特徴分析
3. 学会等名 第二言語習得研究会第31回全国大会
4. 発表年 2020年

1. 発表者名 李在鎬
2. 発表標題 テキストマイニングに基づく学習者コーパス 研究の理論と方法
3. 学会等名 JACET関東支部講演会（招待講演）
4. 発表年 2020年

1. 発表者名 伊集院郁子・李在鎬・青木優子・長谷部陽一郎・村田裕美子
2. 発表標題 複数のコーパス分析に基づく接続詞使用と作文トピックの関係性
3. 学会等名 第22回専門日本語教育学会
4. 発表年 2020年

1. 発表者名 李在鎬
2. 発表標題 ICTを利用した読解授業について
3. 学会等名 第32回日本語教育連絡会議
4. 発表年 2019年

1. 発表者名 李在鎬, 長谷部陽一郎
2. 発表標題 Implementation of Data Visualization Features for Japanese Writing Support System “jWriter”
3. 学会等名 CASTEL/J (国際学会)
4. 発表年 2019年

1. 発表者名 李在鎬
2. 発表標題 教科書に対する読みやすさの評価 jReadability と BCCWJ を用いて
3. 学会等名 母語・継承語・バイリンガル教育 (MHB) 学会2019年度研究大会
4. 発表年 2019年

1. 発表者名 三谷彩, 滝島雅子, 李在鎬
2. 発表標題 学術論文の構成要素に関する計量的分析 日本語教育学分野の研究論文を対象に
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

〔図書〕 計6件

1. 著者名 李在鎬	4. 発行年 2020年
2. 出版社 ひつじ書房	5. 総ページ数 420
3. 書名 講座 言語研究の革新と継承4 認知言語学I「用法基盤モデル」	

1. 著者名 李在鎬	4. 発行年 2020年
2. 出版社 くろしお出版	5. 総ページ数 204
3. 書名 自動詞と他動詞の教え方を考える (学習者作文と自他の使用頻度)	

1. 著者名 牧野成一, 奥野由紀子, 李在鎬, 鎌田修, 嶋田和子, 三浦謙一	4. 発行年 2020年
2. 出版社 凡人社	5. 総ページ数 264
3. 書名 OPIによる会話能力の評価 テスティング、教育、評価に生かす	

1. 著者名 迫田 久美子, 石川 慎一郎, 李 在鎬	4. 発行年 2020年
2. 出版社 くろしお出版	5. 総ページ数 268
3. 書名 日本語学習者コースI-JAS入門	

1. 著者名 池上嘉彦, 山梨正明	4. 発行年 2020年
2. 出版社 ひつじ書房	5. 総ページ数 420
3. 書名 認知言語学 (第6章用法基盤モデル)	

1. 著者名 當作 靖彦, 李 在鎬	4. 発行年 2019年
2. 出版社 ひつじ書房	5. 総ページ数 304
3. 書名 ICT×日本語教育	

〔産業財産権〕

〔その他〕

jreadability日本語文章難易度判別システム  
<https://jreadability.net/>  
 jWriter (学習者作文評価システム)  
<https://jreadability.net/jwriter/>

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	伊集院 郁子  (Ijuin Ikuko)  (20436661)	東京外国語大学・大学院国際日本学研究院・教授   (12603)	論理性の判定
研究分担者	長谷部 陽一郎  (Hasebe Yoichiro)  (90353135)	同志社大学・グローバル・コミュニケーション学部・教授   (34310)	システム開発
研究分担者	青木 優子  (Aoki Yuko)  (90724691)	東京福祉大学・社会福祉学部・講師   (32304)	データ収集

	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	村田 裕美子  (Murata Yumiko)	ミュンヘン大学日本研究センター	データ収集

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
---------	---------