

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 2日現在

機関番号：34310

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500236

研究課題名（和文） 文字列データ解析システムの構築と平安朝文学の伝本と表現に関する総合的研究

研究課題名（英文） Construction of the character string analysis system and the research of the literature of the Heian period about a text and expression.

研究代表者

福田 智子（FUKUDA TOMOKO）

同志社大学・文化情報学部・准教授

研究者番号：50363388

研究成果の概要（和文）：『古今和歌六帖』と『源氏物語』を研究対象として校本システムを開発し、平安朝文学の伝本と表現に関する考察をおこなった。和歌用デジタル校本システムについては、伝本の墨付きの現状を、より論理的に表記するタグ付け規則を案出した。また、散文用校本作成支援ツールは、『源氏物語』の伝本4本のデータ処理をほぼ完了し、計算機を用いた異文箇所の数値化、および Splits Tree による本文系統の視覚化といった、本文異同を把握する一連の手法を確立した。

研究成果の概要（英文）：In this research, by the system which creates the list which showed the difference of the text that we had originally devised, we performed research on a text and expression about the literary works of the Heian period. Those literary works are the anthology of waka poems called "Kokin-Waka-Rokujo" and "Tale of Genji". In the system for WAKA, although the text for analysis needed to be tagged, we devised the more logical rule of tagging. In the system for a prose, processing was mostly completed about four texts of the "Tale of Genji." Moreover, we established a series of methods of recognizing a different partial character string between texts. The method makes a different partial character string a numerical value first using a computer, next classifies a text using Splits Tree, and is visualized.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	900,000	270,000	1,170,000
2012年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学、図書館情報学・人文社会情報学

キーワード：文学情報・和歌・源氏物語・古今和歌六帖・本文異同・情報科学・データマイニング・文字列解析

1. 研究開始当初の背景

国文学研究における計算機利用は、まず、国文学研究資料館において、組織的に提唱されたと言えるだろう。なかでも伊藤鉄也氏に

は、早くも1980年代から、計算機を用いたデータベース作成に関する論考がある。伊井春樹氏、小松茂美氏とともに、1984年から16年の歳月を費やして完成させた『源氏物語

別本集成』(おうふう)は、代表的な成果のひとつである。また、伊井春樹氏編、角川古典大観『源氏物語』CD-ROM(1999年10月)も利用価値が高い。

また、文学作品のCD-ROM版として、おそらくもっとも画期的だったのは、1996年8月に出版された、『新編国歌大観』CD-ROM版(角川書店)であろう。これは、奈良時代から江戸時代までの古典和歌、約45万首を収めた10巻20冊の本を、1枚のCD-ROMに収めたものである。その後、Ver.2(角川学芸出版)が2003年6月に出され、さらに緻密なテキスト検索が可能になった。また、2008年12月に出版された『新編私家集大成』CD-ROM版(エムワイ企画)も、待望されてきたものである。

そして、先の伊藤氏と同じく国文学研究資料館の中村康夫氏も、国文学研究資料館データベース古典コレクション(岩波書店)を、1999年7月より刊行し続けている。さらに、安道百合子氏との共著、『文系のための情報処理入門—パソコンを活用して研究を進めよう—』(和泉書院、2008年7月)では、データベース作成の他、「類歌検索」や「校本作成」などのプログラミング技術を紹介している。

また、電子テキストをCD-ROMのインターフェイス越しに検索するのではなく、計算機プログラムによって解析する研究として、近藤泰弘氏・近藤みゆき氏によるnグラム統計処理を用いた文字列解析が挙げられる。

このように、国文学研究における計算機技術の導入は、近年、急激に盛んになってきた。だがその一方で、国文学研究者のニーズに、コンピュータ技術が適切に対応しているかといえば、必ずしもそうとは言い切れない。国文学研究者自身がプログラミング技術を習得し、実際に研究に生かすことができるのは、ごく稀なケースに限られる。また、市販されている各種CD-ROMでは、検索機能はかなり充実しているが、校本システムは未だ前例がなく、情報科学分野において近年発展の著しいデータマイニング技術も駆使されているとは言い難い。

そこで本研究では、まず、国文学研究者と情報科学研究者とが連携した研究体制を整え、任意に底本を変えて表示できる、全く新しい発想の校本システムを開発する。これを用いることで、従来、人手では事実上不可能に近かった大量で複雑な本文(テキスト)を比較・分類し、平安朝文学の伝本と表現に関する新たな知見を得る。

2. 研究の目的

(1)文字列データ解析システム(デジタル校本システム)の構築

コンピュータの画面上に校本を表示する

ことで、容易に底本が変更でき、本文異同も視認しやすく、また、伝本を逐一増補できる校本システムを構築する。

①和歌用デジタル校本システム

校本画面の表示の他、和歌本文は、原態のままの表記でも、清音仮名・歴史的仮名遣いに改めた本文でも検索を可能にする。傍書によって生じる和歌本文は、プログラムによって自動生成する。見せ消しや補入記号など、あらゆる墨付の検索に対応した、独自のタグ付け規則を案出する。

②物語用校本作成支援システム

①の和歌用システムとは異なり、計算機によって完全な校本を作り上げようとするのではなく、校本作成を支援するシステム(最終段階の仕上げは人間の手に委ねる)の開発を行う。人手で大量のデータを扱うことになるので、ユーザーの使い勝手のよさを考え、インターフェイスを工夫する。

(2)システムを用いた伝本・表現に関する平安朝文学研究

上記2種類の文字列データ解析システムを用いて、平安朝文学のうち、和歌と物語に関する伝本と表現の研究を行う。すなわち、これらのシステムの使用により、諸本の本文異同を把握し、作者(编者)自筆本により近い写本を特定する手掛かりを得るとともに、享受史の観点から、後世の流布本の性格を把握し、異同が生じた理由を探る。また、諸本本文の性格を総合的に整理するための足掛かりをつくる。

和歌用デジタル校本システムを用いる研究対象として、まず、十世紀後半成立かといわれる日本初の類題和歌集『古今和歌六帖』(約4,500首)を採り上げる。有力な古写本の現存が確認されない歌集であり、また、大部でもあるため、総合的な伝本・表現の研究は、これまで必ずしも容易ではなかった。この歌集については、科学研究費補助金基盤研究(C)「文字列データ解析システムの構築と平安中期歌語生成に関する研究」(課題番号19500217、平成19~21年度)からすでに研究に取り組んでおり、本研究でも継続して、電子テキストの作成・増補とシステムの開発をおこない、実際に文学研究に利用する。

また、平安中期、十世紀後半成立の初期百首のうち、『曾根好忠集』所載(順百首)を採り上げる。比較的取り扱いやすい百首という歌数と、主要古写本が三系統存するという点で、システムを考案する上でも、研究価値は高い。

これらの歌集の成立は、王朝女流文学が隆盛を誇る一条朝の前夜にあたり、和歌文学研究において、従来、必ずしも多くの研究の手が及んでいた時期とは言い難い。それだけに、校本作成の価値はじゅうぶんに認められる。

物語用校本作成支援システムを用いる研究対象としては、『源氏物語』を採り上げる。『CD-ROM 角川古典大観 源氏物語』には、大島本・陽明文庫本・保坂本・尾州家河内本の翻字テキストが収録されているが、これら4本の校本を、底本を変えて作成する。飯島本などの新出本については、電子テキストを作成しながら、随時増補していく。『源氏物語』の伝本研究、とくに別本を視野にいたした研究に資するところ大であろう。

3. 研究の方法

本研究を遂行する構成員は、連携研究者6名の他に、研究協力者4名を加える。〈和歌〉〈物語〉の各研究チームが並行して、情報科学研究者がシステムの開発をおこない、文学研究者がその利用による文学研究を行う。

〈和歌〉〈物語〉チームはそれぞれ、同志社大学文化情報学部の実験系授業、および学生研究会を下部組織とする。そこで作成・整備した電子テキストは、研究代表者(福田)が取りまとめ、〈和歌〉〈物語〉チーム(文学研究者)の最終チェックを経る。できあがった電子テキストを対象に、情報科学研究者の開発した校本システムを用いて校本を作成し、伝本研究を行うという手順をとる。

なお、〈和歌〉のデータに関しては、一部、黒田彰子氏(愛知文教大学)から電子テキストの提供を受ける。

4. 研究成果

(1)文字列データ解析システム(デジタル校本システム)の開発

①和歌用デジタル校本システム

本システムでは、先の科研費研究の成果をもとに、文献の紙面に見られる墨付をはじめとする様々な現象を合理的に記号化するため、タグ付け規則を整備するとともに、『古今和歌六帖』の伝本をさらに4本(書陵部蔵御所本・島原図書館蔵肥前島原松平文庫本・神宮文庫蔵林崎文庫旧蔵本、同宮崎文庫旧蔵本)作成し、合計10本の伝本を対象に本文研究を行う準備を整えた。システムの開発は竹田正幸氏に依頼し、また、テキスト作成・本文チェックには、とくに南里一郎氏の協力を得た。

電子テキストを作成する際、入力作業は基本的に人手に頼ることになるため、先の科研費研究当初は、作業の負担軽減のため、ごくシンプルなデータ構造とタグにとどめていた。

■データの構成(旧)

- [1行目] 歌番号(『新編国歌大観』に依拠)
- [2行目] 題・詞書,題・詞書の書き入れ
- [3行目] 作者名,作者名の書き入れ
- [4行目] 歌本文A(原態のままの表記)

[5行目] 歌本文B(清音仮名・歴史的仮名遣いに改めた本文)

[6行目] 左注,左注の書き入れ

[7行目] 歌本文Aの書き入れ

【例:永青文庫本530番歌のデータ】

=====

530

*

*

さよふけていつらん月を/たかやまのみねによそへて/なかめつるかな

さよふけていつらむつきを/たかやまのみねによそへて/なかめつるかな

*

4[3]=5[4|7]r{のしらくも/かくしてしかなイ|のしらくも/かくしてしかな}(2[4-8|4-7]r{てんかもイ|てむかも},2[*|4-7]D{つるかなイ|つるかな})

=====

だが、本研究においては、主要伝本の電子テキストが出揃ったのを機に、上記の形式のデータを計算機により一括変換し、より視認性が高く、かつ合理的な構造に切り替えた。

■データの構成(新)

伝本: 伝本の名称

番号: 歌番号(『新編国歌大観』に依拠)

題詞: 歌題や詞書

→題詞についての書き入れ

作者: 作者名

→作者名についての書き入れ

和歌: 歌本文A(原態のままの表記)

→歌本文B(清音仮名・歴史的仮名遣いに改めた本文)

→歌本文Aについての書き入れ

左注: 左注

→左注についての書き入れ

【例:永青文庫本530番歌の新形式のデータ】

=====

伝本: 永青

番号: 530

題詞: *

作者: *

和歌: さよふけていつらん月を/たかやまのみねによそへて/なかめつるかな

→さよふけていつらむつきを/たかやまのみねによそへて/なかめつるかな

→¥右傍書{4.3-5.7}{のしらくも/かくしてしかなイ}¥REP{4.3-5.7}{のしらくも/かくしてしかな}

→→¥右傍書{2.4-2.8}{てんかもイ}¥REP{2.4-2.7}{てむかも}

→→¥行末注{つるかなイ}¥REP{2.4-2.7}{つるかな}

左注：＊
=====

旧構造では、行数によりデータ項目を規定していたが、新構造では、行頭にデータ項目を明示し、さらに、書き入れの処理はすべて、当該項目の直後に「→」で記すようにした。これにより、従来のタグ付けでは表示し難かった複雑な現象も、比較的自在に表示できるようになった（上記の例は、和歌の右傍書に対して、右傍書と行末注の二つの書き入れが付く場合を示す）。この観点は、複雑な現象を把握するための情報整理の一方法として、他の分野にも利用可能であろう。

このシステムの表示画像の一部を以下に示す。

【画像1】底本（画像では桂宮本を選択）に対する語句の本文異同をもつ歌の歌数表示（表記の異同も表示可能）。

桂宮	永青	寛文	件数
×	×		470
○	×		956
×	○		284
○	○		2784

【画像2】同じ歌番号の歌（自動生成した書き入れ本文も含む）の本文を列举し、異同箇所を句ごとに表示

桂宮 530	永青 530	寛文 530
さよふけて いつらむつきを たかやまの あねにこそへて なかつるかな	さよふけて いつらむつきを たかやまの あねのしるも かくしてしかな	さよふけて いつらむつきを たかやまの あねのしるも かくしてしかな

今後、比較する伝本数を増やしていくことは、システムの機能上は可能である。だが、結果の表示方法には、さらなる工夫が必要であろう。場合によっては、いったん出力したデータを、さらに整理・表示するという方向性も模索すべきであると考えます。

②物語用校本作成支援システム

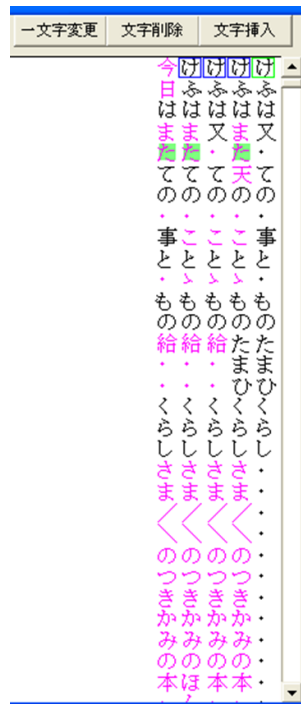
本システムは、坂田桂一氏によって作成されたものである。大量で複雑な本文データを比較・分類するために、計算機で行える処理は計算機に任せ、最小限、人手による作業を

加えるというコンセプトによる、校本作成のための「支援ツール」である。

また、本文データは、各帖をいくつかの節に分け、その節ごとに、本文を1本の文字列として扱う。文節や形態素に分割していないのも、人間の労力を最小限に留めるという考え方による。

とくに、語や文節が、伝本間で前後している場合、テキストを読み込んだ段階では、校本を完全に自動作成することはできないが、ソフト上での修正は、比較的容易である。

【画像3】テキスト表示画面



（横並びの文字の関連付けを崩すことなく、「一文字変更」「文字削除」「文字挿入」をはじめ、文字列をずらすことも可能。必要に応じて、横並びの文字の関連付けを解除することもできる。）

本システムの特徴としては、まず、底本を自由に選択できることが挙げられよう。従来の紙媒体に印刷された校本は、容易に底本を変更することができなかったが、本システムは、校本に用いたすべての本を底本とする校本が、一度に出力できる。どの本を底本とするかの判断は、本文研究上、重要であるが、すべての校本に目を通してから、研究目的に適した校本を選ぶことが可能となった。

また、いったん完成した校本に、新たな伝本を追加することも可能である。電子テキストを新たに作成した際に、これまで作成してきた校本データを最大限利用することができるという利点は大きい。現時点でも、伝本20本の処理が可能である。今後、それ以上の伝本数を処理しようとするならば、システムの機能そのものは拡張し得るが、本文表示の方法については、一考の余地がある。これは、先の和歌用デジタル校本システムと共通する問題である。

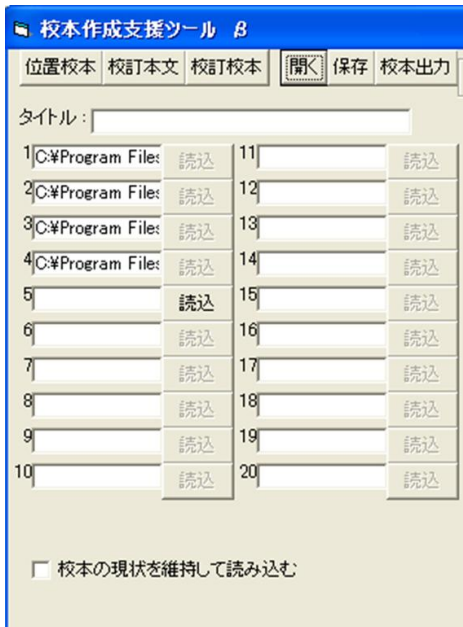
【画像 4】校本出力結果表示（部分）

```

1 | けふは又・ての・事と・ものたまひくらし・さま／＼
2 | ーまた天・こー　　給・　　さま／＼
3 | ー　　・　　・　　・　　給・　　さま／＼
4 | ーまたー　　・　　・　　給・　　さま／＼
5 | 今日ーまたー　　・　　・　　給・　　さま／＼
6 | □|
7 | □|
8 | けふはまた天の・こと、ものたまひくらしさま／＼
9 | ー又・てー　　・　　・　　・　　・　　・　　・　　・　　・
10 | ー又・てー　　・　　・　　・　　・　　・　　・　　・　　・
11 | ーたてー　　・　　・　　・　　・　　・　　・　　・　　・
12 | 今日ーたてー　　・　　・　　・　　・　　・　　・　　・　　・
13 | □|
14 | □|
15 | けふは又・ての・こと、もの給・くらしさま／＼
16 | ー　　・　　・　　・　　・　　・　　・　　・　　・　　・　　・
17 | ーまた天ー　　・　　・　　・　　・　　・　　・　　・　　・
18 | ーまたー　　・　　・　　・　　・　　・　　・　　・　　・　　・
19 | 今日ーまたー　　・　　・　　・　　・　　・　　・　　・　　・

```

【画像 5】テキスト読み込み画面



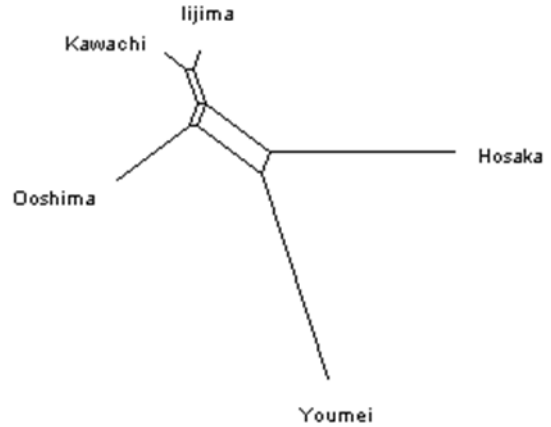
なお、[画像 4] の校本は、もちろんそのまま伝本研究に使用可能であるが、全体の本文異同を一度に把握するのは、やはり難しい。そこで、この出力データを用いて、エクセルのマクロを利用し、異同箇所のカテゴリ・数値化をおこなった（プログラムは深川大路氏による）。

【画像 6】異同箇所の分類・数値化画面

	A	B	C	D	E	F	G
1	0lijima	前斎	宮の御	まいりの	事・中宮		
2	0hosaka						の
3	0kawachi					こと	
4	0ooshima						の
5	0lyoumei	せんさい		うち			の
6							
7	0lijima	1	1	1	1	1	1
8	0hosaka	1	1	1	1	1	2
9	0kawachi	1	1	1	1	2	1
10	0ooshima	1	1	1	1	1	2
11	0lyoumei	2	1	2	1	1	2

すなわち、異同箇所について、底本の本文を「1」とした時、異文は「2」、そして、「1」「2」とも異なる本文があれば「3」……と、自動的に数値化していくのである。このように、すべての異同が数値化されれば、本文全体の異同の状況は、Splits Tree を利用した視覚化によって、より明瞭に示すことができる（矢野環氏の提案による）。

【画像 7】Splits Tree を用いた系統分類例



このようにして、校本作成支援ツールを中心に、エクセルのマクロや Splits Tree を組み合わせることで、ともすれば複雑に絡み合う本文異同を、客観的に把握することができる。

(2) システムを用いた伝本・表現に関する平安朝文学研究

和歌研究においては、『古今和歌六帖』について、出典未詳歌注釈をはじめ、表現研究の一端を研究論文として発表した。また、筑紫平安文学会において、『順百首』の注釈作業をおこない、伝本と表現に関する解説原稿を執筆した。

また、物語研究においては、『源氏物語』の伝本と表現の研究をおこない、その成果は目下、研究論文として準備中である。また、宮崎裕子氏により、開発したシステムが『石清水物語』の伝本研究に用いられ、その有用性が裏付けられた。

詳細については、次項を参照されたい。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 14 件）

- ① 福田智子、〈順百首〉の表現受容—歌合を中心に—、『順百首全釈』、査読有、歌合・定数歌全釈叢書 18、2013、pp. 289

—297

- ② 南里一郎、『曾禰好忠集』所収〈順百首〉の本文について、『順百首全釈』、査読有、歌合・定数歌全釈叢書 18、2013、pp. 229—248
- ③ 福田智子、後世の歌集に採られた〈順百首〉歌の作者認定について一、『順百首全釈』、査読有、歌合・定数歌全釈叢書 18、2013、pp. 298—315
- ④ 福田智子、題と本文の間—『古今和歌六帖』諸本の本文異同と『万葉集』一、同志社国文学、査読有、第 78 号、2013、pp. 16—27
- ⑤ 福田智子、『古今和歌六帖』出典未詳歌注釈稿—第六帖(6) くだに～蓮一、文化情報学、査読有、Vol. 8, No. 2、2013、pp. 一(70)—一七(54)
- ⑥ 福田智子、『古今和歌六帖』本文の変容と享受—伊勢の「雪中の筍」題の屏風歌をめぐる一、文化情報学、査読有、Vol. 8, No. 1、2012、pp. 一(52)—一〇(43)
- ⑦ 福田智子、『古今和歌六帖』出典未詳歌注釈稿—第六帖(8) 尊～ことなし草一、女子大國文(京都女子大学)、査読有、第 151 号、2012、pp. 40—76
- ⑧ 福田智子(他 4 名)、『古今和歌六帖』出典未詳歌注釈稿—第六帖(5) 菊～紫苑一、文化情報学、査読有、Vol. 7, No. 2、2012、pp. 一(70)—一七(54)
- ⑨ 宮崎裕子、『石清水物語』第三系統諸伝本に関する研究(一)、文献探究、査読無、第 50 号、2012、pp. 34—60
- ⑩ 福田智子、平成 21 年 国語国文学界の動向 中古韻文 伝本・歌人・表現—平安中期を中心に— 附 和歌データベースの将来、文学・語学、査読有、第 201 号、2011、pp. 72—75
- ⑪ 福田智子(他 3 名)、『古今和歌六帖』出典未詳歌注釈稿—第六帖(2) 下草～雑の草一、文化情報学、査読有、Vol. 7, No. 1、2011、pp. 82(一)—69(一四)
- ⑫ 福田智子、〈順百首〉の表現摂取—先行歌集・歌合との関わりと『古今和歌六帖』一、文化情報学、査読有、Vol. 6, No. 1、2011、pp. 124(一)—112(一三)
- ⑬ 藤井翔太、『古今和歌六帖』の本文に関する一考察—出典歌集の配列との関わりから—、文化情報学、査読有、Vol. 6, No. 1、2011、pp. 111(一四)—102(二三)
- ⑭ 福田智子(他 3 名) 『古今和歌六帖』出典未詳歌注釈稿—第六帖(4) 薄・篠薄・萩・蘭—(「文化情報学」Vol. 6, No. 1、平成 23(2011) 年 3 月、pp. 101(二四)—84(四一))

[学会発表] (計 2 件)

- ① 福田智子・竹田正幸・南里一郎、『古今和

歌六帖』の本文と表現—デジタル校本の試み—、第 16 回公開シンポジウム「人文科学とデータベース」、2010 年 11 月 27 日、花園大学拈花館 202

- ② 福田智子、『古今和歌六帖』の写本・版中間の本文異同について—万葉歌を中心に—、中古文学会秋季大会、2010 年 10 月 3 日、立命館大学衣笠キャンパス

6. 研究組織

(1) 研究代表者

福田 智子 (FUKUDA TOMOKO)
同志社大学・文化情報学部・准教授
研究者番号：50363388

(2) 研究分担者

なし

(3) 連携研究者

矢野 環 (YANO TAMAKI)
同志社大学・文化情報学部・教授
研究者番号：10111410

田坂 憲二 (TASAKA KENJI)
群馬県立女子大学・文学部・教授
研究者番号：70136406

岩坪 健 (IWATUBO TAKESHI)
同志社大学・文学部・教授
研究者番号：00211764

黒木 香 (KUROKI KAORI)
活水女子大学・文学部・准教授
研究者番号：70195577

竹田 正幸 (TAKEDA MASAYUKI)
九州大学・システム情報科学研究院・教授
研究者番号：50216909

深川 大路 (FUKAGAWA DAIJI)
同志社大学・文化情報学部・助教
研究者番号：10442518

波多野 賢治 (HATANO KENJI)
同志社大学・文化情報学部・准教授
研究者番号：80314532

(4) 研究協力者

南里 一郎 (NANRI ICHIRO)
立命館大学・情報理工学部・非常勤講師
研究者番号：10312191

宮崎 裕子 (MIYAZAKI YUKO)
九州産業大学・国際文化学部・講師
研究者番号：40581533

坂田 桂一 (SAKATA KEIICHI)
同志社大学・法学研究科・博士後期課程
藤井 翔太 (FUJII SHOTA)
同志社大学・文化情報学部・卒業生