

2025 年度中央大学共同プロジェクト 研究実績報告書

1. 概要

研究代表者		所属機関	国際情報学部	2025 年度助成額	
		氏名	橋本 健広	4,521,000 円	
		NAME	Takehiro Hashimoto		
研究 課題名	和 文	文学を科学する：機械学習を用いて文脈にもとづくテキスト間の影響を調べる学際的研究		研究 期間	2024～2025 年度
	英 文	The science of literature: Interdisciplinary study of textual influence based on context-oriented machine learning			

2. 研究組織

※所属機関・部局・職名は 2026 年 3 月 31 日時点のものです。

	研究代表者及び研究分担者		役割分担	備考
	氏名	所属機関/部局/職		
1	橋本 健広	中央大学・国際情報学部・教授	研究全般	
2	小花 聖輝	中央大学・国際情報学部・准教授	機械学習モデルの構築	
3	笹川 浩	中央大学・商学部・教授	テキスト分析（後の時代の影響）	
4	矢島 壮平	中央大学・国際情報学部・准教授	テキスト分析（多文化の影響）	
5	山西 博之	中央大学・理工学部・教授	機械学習モデルの言語学的応用方法	
6	木谷 巖	帝京大学・教育学部・教授	テキスト分析（同時代の新規の影響）	
7	宮川 創	筑波大学・人文社会系・准教授	テキスト分析	
8	古屋 耕平	青山学院大学・経済学部・教授	テキスト分析	
9	佐藤 健	神戸大学・大学教育推進機構 国際コミュニケーションセンター・教授	応用言語学	
合計 9 名				

3. 2025年度の研究活動報告 ※行が不足する場合は、適宜、行を追加してご記入ください。

(和文)

本研究プロジェクトの目的は、1. 特定の文学作品および研究成果を利用することで、イギリス文学研究に特化した機械学習のモデルおよび適用式を構築し、2. 機械学習を用いて意味や文脈にもとづく影響の析出を試みるテキスト分析を行うこと、また、3. 実証研究の機会を増やすことで、文学の研究領域の拡大と、文学研究の再興を目指し、4. 大学をデジタル・ヒューマニティーズ（以下 DH）の人的ネットワーク拠点とし、イギリス文学、DH、自然言語処理の学会、研究所、研究者等と連携を試みることである。

具体的には、以下に示すように①機械学習のモデルおよび適用式の作成と、応用としてのテキスト分析である。

①（機械学習のモデルおよび適用式の構築）

1) イギリス文学作品の韻文を中心とするテキスト群と意味が確定されたコンコーダンスを用いた、イギリス文学に特化した機械学習モデルの構築

2) 文学テキストの分析に適した機械学習モデルの分析方法（適用式）の検討

②（イギリス文学分野における機械学習を利用したテキスト分析）

1. 研究計画の進行状況（研究内容）について

2025年度の活動は、2025年4月の検討会、2026年1月に行った報告会および国際シンポジウム[38]、毎月開催する研究会（人文学テキストの分析研究、デジタル人文学の研究者を交えた発表及びディスカッション）[40]を通して行われた。4月の検討会では今年度の方針について話し合い、6月のシンポジウム（NII 受託研究と共同主催）ではデジタル人文学研究における著作権について小向太郎氏（中央大学）より講演を受けた後、LLMを使用したデジタル人文学研究の発表（研究分担者である古屋耕平、宮川創含む）及びディスカッションを行った[11][20][27][35]。研究期間中に LLM を用いたデジタル人文学に関する2件のワークショップを行った[36][37]。報告会として、1月の国際シンポジウム（Digital Approach to Literary Analysis 2025, 中央大学にて開催）では、本研究プロジェクトの紹介をするとともに、ステファニー・サンチ氏（チューリッヒ大学）の基調講演を含む合計3件のデジタル人文学のテキスト分析の講演を行った[17][38]。

本年度の研究は三方面に分かれる。ひとつは前年度までに作成したデータセットの構築の続きである。ふたつ目は影響の分析を行うために文学テキストの注を集めデータセットを作成していたが、作業に非常に時間がかかるため、注収集のための支援ツールを作成することにしたことである。支援ツールをもとに共同研究者間で影響のデータセット作成の続きを行う予定である。みつつ目は現時点で所有するデータセットから言語モデルを作成し、テキスト分析を行う研究である。以下本プロジェクトの研究計画に即して進行状況を示す。

① 機械学習のモデルおよび適用式の構築 コンコーダンスのデータ化、文学テキストの注の収集、言語モデルの作成、適用式の検討と発表を行った。

1) コンコーダンス 2024年度より行っているシェリーのコンコーダンスのデータセット化の続きを行った。このデータセットについては作成が終了した[3][41]。

1)-2 研究を続ける中で、影響の分析を行うために文学テキストの注を用いてデータセットを構築し、モデルを作成すると有益とわかった。このため、文学テキストの注を2024年度より収集している。中の収集には大変な手間と労力がかかるため、一旦中止し、注を収集しやすくするための支援ツールを本年度は作成した[33]。

1)-3 モデルの作成 ファインチューニング用のデータを順次大規模言語モデル(BERT、OpenAI、Claude など)にファインチューニングさせている。Paraphrase-xlmr や paraphrase-mpnet といった言語モデルをベースに、本研究で収集したデータセット（影響のデータセット440件、比喩のデータセット780件ほど）を用いたファインチューニング

済言語モデルを作成した[34]。データセットの収集は今後も続ける予定である。

2) 適用式の作成 英文学に特化したモデルと文学研究を接続できるような計算方法を考案中である。2024 年度に検討した内容をもとに適用式を含めた発表を行った[1][4][18]。

② 機械学習を利用したテキスト分析 現時点でのモデルを元に学会や研究会で発表を行った[2][15][16][19][28][30][40][42]。影響のデータセットをファインチューニングさせたモデルのほうがファインチューニングなしのモデルより、影響分析が向上することがわかっている。この他、本研究プロジェクトで得た大規模言語モデルや機械学習の知見を元に、影響分析につながる研究を学会や研究会等で発表している[5-9][10][12-14][21-26][42]。研究の一つは『週刊読書人』に掲載された[30]。

なお、本共同研究プロジェクトを含めた研究活動を行う形で提案した課題が、国立国語研究所「研究データエコシステム構築事業」のユースケース創出課題として採択された[29]（総額 600 万円ほど）。同じく申請した科研費 B は残念ながら採択に至らなかったが、別の公募として Schmidt Science の研究費に応募した。

この他、毎月主催する研究会[40]や各学会等での交流および海外研究者の招聘[39]により、デジタル人文学（デジタル・ヒューマニティーズ）関連の数多くの研究者との研究交流及び連携を行った。これにより「4. 大学をデジタル・ヒューマニティーズ（以下 DH）の人的ネットワーク拠点とし、イギリス文学、DH、自然言語処理の学会、研究所、研究者等と連携」という目標を達成できた。

計画通り遂行されていない場合の対応状況

本研究プロジェクトではデータセットの作成と、大規模言語モデル及び人文学テキストの分析を同時並行する計画であったが、データセットの作成に当初想定した以上に時間がかかっている。特に、コンコーダンスのデータセットは作成できたものの、分析に大きく影響を与えると考えられる影響のデータセットの作成に時間がかかっている。これは研究者が文学テキストの注を一つ一つ確認し、影響あるとされるテキストを手作業で集めているためである。そのため 2025 年度は、注の収集を一旦停止し、作業の一部を機械化する支援ツールを作成することにした[33]。ある程度の形ができたため、今後は支援ツールをもとに、影響のデータセットの作成を続ける予定である。

次年度以降に残された課題

次年度以降の課題として、データセットの拡充、大規模言語モデルの精度の向上、研究者交流の維持があげられる。2025 年度は影響のデータセットの作成を一旦停止し、注収集のための支援ツールの作成を行った[33]。注収集の支援ツールは年度末までに動作可能なところまでできたため、2026 年度以降は研究分担者と文学テキストの注の収集を続け、2000 件の影響のデータセットの作成を目指したい。データセットの数を増やすことにより、大規模言語モデルの精度の向上が見込まれる。

今後は文学の研究分担者とともに注を収集し、情報の研究分担者とともに言語モデルをファインチューニングし、できたモデルを元に分析を行い、すべての研究分担者を交えて比較し評価する予定である。その結果は、研究期間終了後 2 年以内に研究成果として、所属機関の研究会等において報告し、共同研究費の助成を受けたことを明記して本学が刊行する機関誌又は学会誌等に発表を行う予定である。

なお、本研究プロジェクトで作成を依頼した注抽出支援ツールについては、2026 年 7 月に国際学会 DH2026（韓国）で発表が予定されている（研究代表者：筒井遥）[43]。

またできあがった機械学習モデルを様々な研究者に使っていただき、文学の研究領域の拡大に貢献したい。そのため、二年間の共同研究プロジェクトで培った研究者交流・連携の場を維

持・拡大に努めたい。

2. 研究費の執行状況

研究費はおおむね予定通りに執行した。データセット作成のための資料や機材、コンコーダンスのデジタル化、大規模言語モデルのプラットフォーム利用料、学会発表などに研究費を使用した。手数料は、当初 Google Cloud Platform の使用料として 100 万円を計上していたが、OpenAI、ChatGPT、Claude、Dify 等、他にも同内容でより簡便に利用できるサービスがあることがわかり、そちらを使用した利用料となる。謝金・謝礼品は、コンコーダンスのデジタル化のために当初パートタイム職員給与での支出を考えていたが、委託報酬の枠で利用する方が簡便とわかりそちらで報酬を支払ったため、パートタイム職員給与が減り謝金・謝礼品が増えている。旅費・交通費については、海外での学会参加に費用が当初想定以上にかかったが、円安や物価高騰を反映した結果と考えられる。またデータセット作成のための文学テキストの資料やデジタル化のための機材、研究発表のための研究書等を購入したため、消耗品の利用が増加した。

3. 研究成果の公表

2025 年度は以下の通り研究成果を公表した。

共同研究プロジェクトについては、6 月および 1 月のシンポジウムで概要を紹介した [11][17][35][38]。構築したデータセットについては、大学紀要に論文を発表し [3]、オンライン上に保存した [41]。

作成した大規模言語モデルについては、橋本と古屋が [2][4][15][16][19][20][28][30][40][42] で論文及び学会発表等を行い、モデル自体についてはオンライン上で公開している [34]。

計算式については、[1][4][18] で論文及び学会発表を行った。その他、関連するデジタル人文学分野における人文テキストの機械学習を用いた分析について、橋本、宮川が [5-9][10][12][21-27] で論文及び研究発表を行っている。また毎月の研究会での発表も行った [40]。

4. 研究分担者の活動

共同研究としての実態

2025 年度も検討会、報告会、ワークショップ、および毎月の研究会を軸に、各自機械学習モデルを用いた人文テキストの分析を行うことで共同研究を実施した。

研究分担者の活動

古屋、宮川は共同研究プロジェクトで作成したモデルやデータセットを用い、テキスト分析の分析を行った。特に、古屋は共同研究プロジェクト参加当初、デジタル人文学の分野は初見であったが、研究会の参加を通して知見を深め、学会発表や論文の掲載に至ったことは特筆すべきである。

2025 年度は、データセットの作成を一旦停止し、データセット作成のための支援ツールに注力した。そのため、笹川、木谷、佐藤は毎月の研究会に参加し文学、英語教育の分野からの議論を行うことで研究に貢献した。

研究成果の公表

古屋、宮川は、毎月の研究会および学会発表や論文掲載を通して、共同研究プロジェクトのデータセットやモデルを用いた研究成果の公表を行った [40][19][20][27][28][42]。

5. 本学における教育活動への還元

2026 年 1 月 19 日に開催した報告会および国際シンポジウム (Digital Approach to Literary Analysis 2025) [38]において、研究者のみならず学部学生、大学院生に対しても告知を行い、

シンポジウムを開催した。本研究プロジェクトの概要およびデジタル人文学における機械学習を用いた各種分析研究を発表し、学生のデジタル人文学に関する理解を深めた。

大規模言語モデルを用いた2件のワークショップについては政策文化研究所と合同で開催することで学内の研究者に対しデジタル人文学の知見の普及に努めた。

(英文)

This research project aims: 1. to build models and application formulas for machine learning specific to English literary studies by using specific literary works and academic resources; 2. to use machine learning to conduct text analysis that attempts to analyze context-based influences; 3. to increase opportunities for empirical research 4. to make the university a human network center for Digital Humanities (DH), and to attempt to collaborate with academic societies, research institutes, and researchers in the fields of English literature, DH, and natural language processing.

Specifically, as shown below, we will (1) create models and application formulas for machine learning and (2) analyze texts as an application of machine learning.

1- Construction of models and application formulas for machine learning

1) Construction of a machine learning model specific to English literature, using a collection of texts, mainly poetry, and concordances that show meanings of the texts.

2) Investigation of analysis methods (application formulas) for machine learning models suitable for analyzing literary texts

2- Text analysis using machine learning in the field of English literature

1. Progress of the Research Plan (Research Content)

We carried out activities in FY2025 through a review meeting in April 2025, a reporting meeting and international symposium held in January 2026[38], and monthly research meetings (analysis of humanities texts, presentations and discussions involving researchers in digital humanities)[40]. At the review meeting in April, we discussed the research plan for this year, and at the symposium in June (co-sponsored with NII commissioned research), We held a lecture on copyright in digital humanities research by Taro Komukai (Chuo University), presentations on digital humanities research using large language model (LLM) (including research collaborators Kohei Furuya and So Miyagawa), and discussions [11][20][27][35]. We also held two workshops on digital humanities using LLM during the research period [36][37]. As a report meeting, we held the international symposium in January (Digital Approach to Literary Analysis 2025, held at Chuo University), and we introduced this research project and presented a total of three lectures on digital humanities text analysis, including a keynote speech by Stephanie Santschi (University of Zurich) [17][38].

This year's research is divided into three areas. First, dataset creation: we continued dataset creation from the previous year. Second, creating a supporting tool: while we were collecting annotations from literary texts to create a dataset for influence analysis, this task proved

very time-consuming, so we decided to create a support tool for annotation collection. We plan to continue creating the influence dataset with our collaborators using this support tool. Third, fine-tuning LLM and text analysis: we created a language model from the dataset we currently possess and conducted text analysis. The progress is shown below in accordance with this project's research plan.

① Construction of machine learning models and application formulas: We digitized concordance data, collected annotations from literary texts, created language models, and examined and presented application formulas.

1) Concordance: We continued the work on creating a dataset of Shelley's concordances, which began in 2024. This dataset is now complete [3][41].

1)-2: Annotation Extraction Tool: During the course of our research, we found it beneficial to construct a dataset using annotations from literary texts and create a model for analyzing influence. Therefore, we began collecting annotations from literary texts in 2024. Because this collection process is extremely time-consuming and labor-intensive, we temporarily suspended it and created a support tool this year to facilitate annotation collection [33].

1)-3: Model Creation: We are sequentially fine-tuning large-scale language models (BERT, OpenAI, Claude, etc.) using data for fine-tuning. Based on language models such as Paraphrase-xlmr and paraphrase-mpnet, we created a fine-tuned language model using the datasets collected in this research (approximately 440 influence datasets and 780 metaphor datasets) [34]. We plan to continue collecting datasets in the future.

2) Development of Application Formulas: We are currently devising a computational method that can connect a model specifically for English literature with literary research. Based on the findings considered in FY2024, we presented the application formulas [1][4][18].

② Text Analysis Using Machine Learning: We presented our current model at academic conferences and workshops [2][15][16][19][28][30][40][42]. We found that models with fine-tuned influence datasets perform better influence analysis than models without fine-tuning. In addition, based on the large-scale language models and machine learning insights gained in this research project, we have presented research related to influence analysis at academic conferences and workshops [5-9][10][12-14][21-26]. One of our studies was published in the weekly magazine "Shukan Dokushojin" [30].

Furthermore, a research project proposed in the form of this joint research project, including other research activities, was selected as a use case creation project for the National

Institute of Informatics' "Developing a Research Data Ecosystem for the Promotion of Data-Driven Science " [29]. Unfortunately, the Grant-in-Aid for Scientific Research (B) that I also applied for was not accepted, so I applied for research funding from Schmidt Science.

In addition, through monthly research meeting we organized [40], exchanges at various academic societies, and an invitation of overseas researcher [39], we engaged in research exchanges and collaborations with numerous researchers related to digital humanities. This allowed us to achieve our goal of "4. to make the university a human network center for Digital Humanities (DH), and to attempt to collaborate with academic societies, research institutes, and researchers in the fields of English literature, DH, and natural language processing."

Response to Delays in Planned Implementation

This research project planned to simultaneously create a dataset and analyze a large-scale language model and humanities texts. However, the creation of the dataset is taking longer than initially anticipated. In particular, while the concordance dataset has been created, the creation of the influence dataset, which is expected to significantly impact the analysis, is taking longer than expected. This is because researchers are manually checking each annotation in literary texts and collecting relevant texts considered to have influence. Therefore, in FY2025, we decided to temporarily suspend the collection of annotations and create a support tool to semi-automate annotation extraction [33]. Now that a certain level of progress has been made, we plan to continue creating the influence dataset based on the support tool.

Challenges Remaining in the Next Year and Beyond

Challenges for the next year and beyond include expanding the dataset, improving the accuracy of the large-scale language model, and maintaining researcher exchange. In FY2025, the creation of the influence dataset was temporarily suspended, and a support tool for annotations collection was created [33]. The support tool for collecting annotations was made operational by the end of the fiscal year. Therefore, from fiscal year 2026 onwards, we will continue collecting annotations from literary texts with the research collaborators, aiming to create a dataset of 2000 influence entries. Increasing the number of datasets is expected to improve the accuracy of the large-scale language model. Furthermore, we hope to contribute to the expansion of the field of literary research by having various researchers use the completed machine learning model. To this end, we will strive to maintain and expand the opportunities for researcher exchange and collaboration cultivated during the two-year joint research project.

2. Status of Research Fund Execution

Research funds were executed largely as planned. Research funds were used for materials

and equipment for dataset creation, digitization of concordances, platform usage fees for the large-scale language model, and conference presentations. Initially, 1 million yen was allocated for Google Cloud Platform usage fees, but we noticed that other services such as OpenAI, ChatGPT, Claude, and Dify offer similar services more conveniently, and the fees for those services were instead used. Initially, we considered using part-time staff salaries to cover payment for the digitization of concordances, but it turned out to be simpler to use the commissioned compensation framework, so we paid the fees that way. As a result, part-time staff salaries decreased, while honoraria and gifts increased. Regarding travel and transportation expenses, attending overseas conferences cost more than initially anticipated, which is thought to be a result of the weak yen and rising prices. In addition, the use of consumables increased because we purchased materials such as literary texts for creating datasets, equipment for digitization, and research books for research presentations.

3. Publication of Research Results

In FY2025, the following research results were published.

Regarding the joint research project, an overview was presented at the symposiums in June and January [11][17][35][38]. Regarding the constructed dataset, we published a paper in the university journal [3] and we stored it online [41].

Regarding the large-scale language model created, Hashimoto and Furuya published papers and presented at academic conferences in [2][4][15][16][19][20][28][30][40][42], and the model itself is publicly available online [34].

Regarding the calculation formula, we published papers and presentations in [1][4][18]. In addition, regarding the analysis of humanities texts using machine learning in related digital humanities fields, Hashimoto and Miyagawa published papers and research presentations in [5-9][10][12][21-27]. We also gave presentations at monthly research meetings [40].

4. Activities of Research Collaborators

Actual Status of the Collaborative Research

In FY2025, the collaborative research was conducted primarily through review meetings, reporting sessions, workshops, and monthly research meetings, with each participant individually analyzing humanities texts using machine learning models.

Activities of Research Collaborators

Furuya and Miyagawa conducted text analysis using models and datasets created in the collaborative research project. In particular, it is noteworthy that Furuya, who initially had no prior knowledge of the field of digital humanities, deepened his understanding through participation in research meetings, leading to presentations at academic conferences and publication of papers.

In FY2025, we temporarily suspended the creation of datasets, and the focus shifted to support tools for dataset creation. Therefore, Sasagawa, Kitani, and Sato contributed to the research by participating in monthly research meetings and engaging in discussions from the fields of literature and English language education.

Publication of Research Results

Furuya and Miyagawa published research results using the collaborative research project's datasets and models through monthly research meetings, academic presentations, and publication of papers [40][19][20][27][28][42].

5. Contribution to Educational Activities at Our University

At the reporting meeting and international symposium (Digital Approach to Literary Analysis 2025) held on January 19, 2026, we welcomed not only researchers but also undergraduate and graduate student. We presented an overview of this research project and various analytical studies using machine learning in digital humanities, deepening students' understanding of digital humanities.

Two workshops using large-scale language models were held jointly with the Institute of Policy and Cultural Studies, aiming to disseminate knowledge of digital humanities to researchers within the university.

4. 主な発表論文等（予定を含む）

2025年度に行った共同研究プロジェクトの研究課題としての成果内容についてご記載ください。

※行が不足する場合は、適宜、行を追加してご記入ください。

【学術論文】（著者名、論文題目、誌名、査読の有無（査読がある場合は必ず査読有りと明記してください）、巻号、頁、発行年月）＜発表予定を含む。但し、投稿中、投稿準備中のものは除く＞

- [1] 橋本健広「文学批評から大規模言語モデルへ—単語埋め込みの組み換えによる文学テキスト解釈の試み」『言語処理学会第31回年次大会発表論文集（2025年3月）』, 1605-1609 頁, 2025年3月, https://www.anlp.jp/proceedings/annual_meeting/2025/pdf_dir/P4-25.pdf.
- [2] 橋本健広「イギリス文学テキストの影響分析に特化したLLMモデルの開発に向けて」『人文科学とコンピュータシンポジウム2025論文集：社会をつなぎ直す情報技術』, 査読有り, 373-380 頁, 2025年12月, <https://ipsj.ixsq.nii.ac.jp/records/2006226>.
- [3] Hashimoto, Takehiro. "The Contemporary Meaning and Its Application of Annotated Concordances." *Japanese Journal of Global Informatics*, vol. 6, pp.23-45, March 2026.
- [4] 橋本健広「BERTによる影響分析：文学テキストの単語埋め込みを考える」『研究報告人文科学とコンピュータ(CH)』, vol. 2026-CH-140, no. 24, pp. 1-5. 2026-01-25, ISSN, 2188-8957. <https://ipsj.ixsq.nii.ac.jp/records/2006797>.

- [5] Miyagawa, So, Eliese-Sophia Lincke, and Heike Behlmer. "Stylometric Analysis of Authorship of the Coptic Homily, On Christian Behaviour." *Coptica* 23, 査読有り, (2025): 13-26.
- [6] Miyagawa, So, Yuki Kyogoku, Yuzuki Tsukagoshi, and Kyoko Amano. "Computational Pathways to Intertextuality of the Ancient Indian Literature: A Multi-Method Analysis of the Maitrāyaṇī and Kāṭhaka Saṃhitās." *Journal of Data Mining and Digital Humanities* 15085, 査読有り, (2025): 1-12.
- [7] Miyagawa, So, Luis Morgado da Costa, Laura Slaughter, and Heike Behlmer. "Automatic Detection of Coptic Text Reuse: Applying Coptic Wordnet to Intertextuality Studies in Selected Coptic Monastic Writings." In *Proceedings of the 13th Global Wordnet Conference*, 査読有り, 179-184. Kota Kinabalu: Global Wordnet Association, 2025.
- [8] Miyagawa, So (宮川創). 「文本重用偵測與互文性網絡：數位人文的計算分析新方法」. 『第十六屆數位典藏與數位人文國際研討會論文集』, 816-822. 2025.
- [9] Miyagawa, So (宮川創), Eliese-Sophia Lincke, and Heike Behlmer. 「スタイロメトリによるコプト語文献の著者帰属の再検討」. 『言語処理学会第 31 回年次大会発表論文集』, 167-171. 2025.

【学会発表】（発表者名、発表題目、学会名、開催地、開催年月）

- [10] 橋本健広「ロマン派時代の戦争と出版物への影響」日本英文学会第 97 回全国大会シンポジウム第 2 部門「デジタル・ヒューマニティーズの発展可能性：英文学研究における現在地」（於東京大学）にて発表および司会, 2025 年 5 月 31 日.
- [11] 橋本健広「プロジェクト紹介」文学研究とプラットフォームキックオフシンポジウム（於中央大学）にて発表および司会, 2025 年 6 月 7 日.
- [12] Hashimoto, Takehiro. "How Does War Affect Romantic Literature? Topic Modeling Romantic Documents." Paper presented at DH2025, Universidade NOVA de Lisboa (Lisbon, Portugal), July 17th, 2025. <https://doi.org/10.5281/zenodo.19494801>.
- [13] 橋本健広「大規模言語モデルを使用した文学研究のための研究資源およびオープンプラットフォームの構築」研究データエコシステム構築事業シンポジウム 2025：AI とデータ基盤がつなぐ、多様な知の最前線（於一橋講堂）パネルディスカッション「人文系データがもたらす分野融合の可能性」（モデレータ：大向 一輝）にて発表, 2025 年 10 月 9 日.
- [14] 橋本健広「大規模言語モデルを使用した文学研究のための研究資源およびオープンプラットフォームの構築」研究データエコシステム構築事業シンポジウム 2025：AI とデータ基盤がつなぐ、多様な知の最前線（於一橋講堂）にて発表, ポスター発表, 2025 年 10 月 9 日.
- [15] 橋本健広「大規模言語モデルの文学研究への適用可能性——影響やインターテクスチュアリティは分析できるか」「DHSS コロキウム「文学」を読む／書く方法：デジタルと世界文学」（代表：須田永遠）主催：名古屋大学デジタル人文社会科学研究推進センター（於名古屋大学）にて発表, 2025 年 12 月 10 日.

- [16] 橋本健広「イギリス文学テキストの影響分析に特化した LLM モデルの開発に向けて」じんもんこ ん 2025（於九州大学伊都キャンパス）にて発表，ポスター発表，2025 年 12 月 14 日．
- [17] Hashimoto, Takehiro “The science of literature: Interdisciplinary study of textual influence based on context-oriented machine learning” in 2025. Project Report presented at Digital Approach to Literary Analysis 2025, Chuo University (Tokyo). January 19th, 2026.
- [18] 橋本健広「BERT による影響分析：文学テキストの単語埋め込みを考える」第 140 回人文科学とコンピュータ研究発表会（於アートホテル石垣島）にて発表，2026 年 2 月 1 日．

- [19] 古屋 耕平「ホーソン作品の原型——SBERT を用いた初期詩 11 篇と The Scarlet Letter の意味的類似性比較」日本ナサニエル・ホーソン協会第 43 回全国大会，シンポジウム「詩人ホーソンの横顔—ホーソンの詩を読む」2025 年 6 月 22 日，専修大学．
- [20] 古屋 耕平「SBERT を用いたテキスト間の意味的類似性比較——英語圏文学研究への応用」文学研究とプラットフォームキックオフシンポジウム，2025 年 6 月 7 日，中央大学市ヶ谷田町キャンパス，

- [21] Miyagawa, So. “Uncovering Old Testament Intertextuality in Coptic Literature: A Computational Analysis using TRACER and Passim.” Paper presented at the 25th Congress of the International Organization for the Study of the Old Testament (IOSOT2025), August 14, 2025.
- [22] Miyagawa, So (宮川創). 「古代末期の修道院文献の間テキスト性ネットワーク：大規模言語モデルによる探知と分析」. Invited paper presented at 英語コーパス学会 (JAECs) 2025 年度・第 51 回大会, September 2, 2025.
- [23] Miyagawa, So. “Advancing Text Reuse Detection in Coptic Literature: A Comparative Study of Large Language Models and Traditional Computational Methods.” Paper presented at Bytes and Bygones - Digital and Computational Analyses of Ancient Cultures, September 16, 2025.

<p>[24] <u>Miyagawa, So</u>. "Intertextuality Studies in the AI Revolution." Invited paper presented at Korean Association for Digital Humanities (KADH) 2025 "Reimagining the Humanities: Big Data, Culture and the Public," October 31, 2025.</p> <p>[25] <u>Miyagawa, So</u>. "Biblical Intertextuality Analysis in Coptic Literature through Large Language Models." Paper presented at the Society of Biblical Literature (SBL) Annual Meeting 2025, November 23, 2025.</p> <p>[26] <u>Miyagawa, So (宮川創)</u>. 「コプト語初期キリスト教修道文献における聖書引用—古典文献学的分析とデジタル人文学的分析の融合—」. Invited paper presented at 同志社大学—神教学際研究センター公開講演会, December 6, 2025.</p> <p>[27] <u>宮川創</u>. 「間テキスト性分析における諸大規模言語モデルとテキスト・リユース探知プログラム TRACER の使用比較」文学研究とプラットフォームキックオフシンポジウム, 2025 年 6 月 7 日, 中央大学市ヶ谷田町キャンパス,</p>	<p>【図 書】(著者名、出版社名、書名、刊行年)</p> <p>[28] <u>古屋 耕平</u>「ハーマン・メルヴィルと十九世紀コレラ流行——移民、都市、伝染病」『病と障害のアメリカルネサンス疫病——ディサビリティ、レジリエンス』高尾直知、伊藤詔子、辻祥子、野崎直之編著、小鳥遊書房、2025 年 5 月、頁 45-68。</p> <p>【その他】(知的財産権、ニュースリリース等)</p> <p>受託研究費</p> <p>[29] <u>橋本健広</u>. 2025 年 4 月～2027 年 3 月 「AI 等の活用を推進する研究データエコシステム構築事業」にかかるユースケース創出課題「大規模言語モデルを使用した文学研究のための研究資源およびオープンプラットフォームの構築」提案代表者</p> <p>PI: MEXT Research Data Ecosystem Project – Resources and LLM-based Analysis on an Open Platform for Literary Studies (2025-2027), National Institute of Informatics, 6400000</p> <p>https://www.nii.ac.jp/creded/creded_result.html</p> <p>ニュースリリース</p> <p>[30] 須田永遠, <u>橋本健広</u>, 日比嘉高, ホイト・ロング, 秋草俊一郎. 「DHSS コロキアム『文学』を読む／書く方法: デジタルと世界文学」レポート」週刊読書人, 2026 年 2 月 13 日, 2 頁.</p> <p>受賞</p> <p>[31] <u>Miyagawa, So (宮川創)</u>. 第 30 回所長賞 (特別所長賞). 国立国語研究所, June 2025.</p> <p>[32] <u>Miyagawa, So (宮川創)</u>. 令和 7 年度若手教員奨励賞. 筑波大学, November 2025.</p>
---	--

ツール・言語モデル

- [33] 筒井遥, 橋本健広. 「抽出くん」. テキストの注を集めるための半機械化ツール. March 2026. <https://www.bunsekisan.jp/chushutsu/>
- [34] Hashimoto, Takehiro. hast2/2025-paraphrase_xlmr_influence-figure_v1. Hugging Face. 2025 年 9 月. [影響と比喩のデータセットを用いてファインチューニングした大規模言語モデル。Paraphrase_xlmr および paraphrase_mpnet を用いて影響のみ、比喩のみ、影響と比喩の計 6 種類]

主催した国際シンポジウム、ワークショップ等

- [35] 文学研究とプラットフォームキックオフシンポジウム主催. 2025 年 6 月 7 日. 中央大学. 小向太郎氏、小川潤氏、古屋耕平氏、宮川創氏を招いてシンポジウムを開催した。
<https://sites.google.com/view/literaryanalysisandplatform202/> [Literary Studies and Platform Kick Off Symposium]
- [36] ワークショップ「文学研究とプラットフォームワークショップ：生成 AI によるプログラミングの実例ーコンコーダンスの作成を例に」2025 年 7 月 8 日. 中央大学. 鎌田明子氏（東京農業大学）、筒井遥氏（東京大学大学院生）による AI プログラミングの共同ワークショップを開催した。
<https://sites.google.com/view/literaryanalysisandplatform203>. [Workshop: Gen AI Programming: Creating a Concordance]
- [37] ワークショップ「文学研究とプラットフォームワークショップ：生成 AI によるプログラミングの実例ー議会資料を例に」2026 年 1 月 16 日. 中央大学. 亀田堯宙氏（人間文化研究機構）、渡部春佳氏（中央大学国際情報学部）による AI プログラミングの共同ワークショップを開催した。
<https://sites.google.com/view/literaryanalysisandplatform204>. [Workshop: Gen AI Programming: Using Parliament Resource]
- [38] 国際シンポジウム Digital Approach to Literary Analysis 2025（人文学におけるテキスト分析のデジタル的試み 2025）主催. 2026 年 1 月 19 日. 中央大学. ステファニー・サンチ氏（チューリッヒ大学）、竹内綾乃氏（国文学研究資料館）、筒井遥氏（人間文化研究機構）を招いて国際シンポジウムを開催した。 <https://sites.google.com/view/chuodala2025/>

研究者招聘

- [39] 研究者交流計画外国人研究者受入れ. 2026 年 4 月 - 2026 年 4 月. 中央大学国際センター2026 年度研究者交流計画外国人研究者受入れ制度を通じ、Simone Rebora 氏（ヴェローナ大学）を受け入れ、研究、交流、学生への教育活動を行った（2026 年 12 月 9 日～12 月 29 日）。

研究会

[40] 毎月第一日曜日に同プロジェクトに関わる人文系テキストのデータ分析の研究会を開催した。

2026/3/6 宮川創先生（筑波大学）間テキスト性分析ツール ICoMa（Intertextuality Collation Machine）の開発

2026/2/6 鄭弯弯先生（名古屋大学）言語データに基づく感情推定—単純極性を越えた感情モデリング—

2026/1/4 2026 年の研究懇談

2025/12/7 三輪玲以佳先生（同志社大学大学院文化情報学研究科 M1）空海伝の発展を捉えるためのテキスト再利用の可能性

2025/11/2 Artem Suslov 先生（北海道大学）Large Language Models in the Generalization of Stylistic Features

2025/10/5 筒井 遥先生（人間文化研究機構）英文学批評の言説地図—雑誌間の主題分布と語彙特性の計量的記述

2025/9/7 宮川創先生（筑波大学）間テキスト性の探知・分析アプリケーションの開発

2025/8/5 橋本健広（中央大学）生成 AI を使用した影響分析

2025/7/6 宮川創先生（筑波大学）間テキスト性分析における諸大規模言語モデルとテキスト・リユース探知プログラム TRACER の使用比較

2025/5/11 古屋耕平先生（青山学院大学）Nathaniel Hawthorne の初期詩 11 篇と The Scarlet Letter の意味的類似性比較

2025/4/6 今年度の活動内容について（検討会）

データセット

[41] Hashimoto, Takehiro. “Dataset for Ellis’s concordance to Shelley’s poetry.” hast2/Ellis_Shelley. September, 2025.

研究発表

[42] 古屋耕平「文系と AI——アメリカ文学の場合」2025 年度東大英文学会総会（於東京大学）にて発表，2026 年 3 月 14 日．

[43] Tsutsui, Haruka, and Takehiro Hashimoto. “A Human-in-the-Loop Tool for Resolving “Dirty” Citations in Multilingual Scholarly and Literary Texts.” Paper will be presented at DH2026, Daejeon Convention Center (Daejeon, South Korea), July, 2026.