



Title	プロジェクトの目的と活動
Author(s)	田畠, 智司
Citation	言語文化共同研究プロジェクト. 2023, 2022, p. 1-4
Version Type	VoR
URL	https://doi.org/10.18910/91698
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

「テクストマイニングとデジタルヒューマニティーズ」 プロジェクトの目的と活動

本共同研究は、自然言語処理、コーパス言語学・計量言語学、数理統計学、データマイニング、機械学習など、諸分野の知見を有機的に統合した方法論を開発し、テクストマイニングを応用して人文学、言語文化学の諸問題にアプローチする、すなわち「デジタルヒューマニティーズ (Digital Humanities)」の実践と理論的精緻化の可能性を探る営みである。このプロジェクトは、2001年度に岩根久教授、緒方典裕助教授、および筆者の3名でスタートした「電子化言語資料分析の方法論」を基礎とするが、2003年度から名称を一部改め、言語文化研究科の大学院生もメンバーに加わった。2006年度には三宅真紀助教の加入を得て、対象言語も英・仏・ギリシャ語に拡がった。2011年には言語文化教育論講座に着任した今尾康裕講師が加入した。2014年度後期から、さらに Hodošček Bor 講師が加わった。そして、2019年度をもって退職された岩根久教授の後任として、2020年度に山田彬堯講師が着任し、言語文化研究科と文学研究科の統合により人文学研究科が設立された2022年には人文学林より菅原裕輝特任助教が加入し、現在の陣容となっている。(職位はいずれも当時)。2016年度から、プロジェクトの名称を、当該リサーチコミュニティの名称としてより相応しい「テクストマイニングとデジタルヒューマニティーズ」にアップデートしたが、研究の系統は創始時より常に一貫している。

「テクストマイニングとデジタルヒューマニティーズ」プロジェクトは大きく分けて二つの層で構成されている。一つは研究基盤となるコーパス、テクストアーカイヴの開発・構築、もう一つは構築したコーパス、テクストアーカイヴからのデータ抽出法研究、並びに得られた高次元の言語データの計量分析である。前者には英・仏語の文学作品や、聖書(共観福音書)などの電子テクスト化、ロシア語政治演説コーパス、近代日本文学コーパスの編纂、マークアップ言語 XML による TEI (Text Encoding Initiative : デジタル化したテクストの国際互換規格の枠組) に準拠したタグ付けなど、人文学資料のデジタル化やマークアップ法、データ符号化方法論の開発などが含まれる。一方、高次元人文学データ分析の事例として、語彙・語法、コロケーション、意味構造、語用論などのレベルにおける言語使用の実態研究、高度な数理モデルや機械学習を応用した言語分析やテクストマイニング、文学作品の言語特徴の特定や、使用域間の言語変異や文体識別問題の考察、著者推定法の精密化研究を挙げることができる。

本プロジェクト班は人文学研究科の専任教員6名と名誉教授1名(今尾康裕、菅原裕輝、田畠智司、Hodošček Bor、三宅真紀、山田彬堯、岩根久名誉教授)、当研究科博士後期課程在学生7名(黄晨斐、岡部未希、徐勤、福本広光、藤田郁、竹森ありさ、涌井萌子)、博士前期課程在学生4名(小堀彩夏、曹芳慧、Vogatza Dimitra、李晨婕)、研究生1名(肖媛媛)に加え、OGの大坂医科大学浅野元子氏(2020年3月博士学位取得)・名古屋外国語大学杉山真央氏(2019年3月博士学位取得)、本学非常勤講師の高橋新氏、南澤佑樹氏(本研究科博士課程修了)、摂南大学後藤一章氏(本研究科博士課程修了)、帝塚山学院大学八野幸子氏(本研究科博士課程修了)、国立国語研究所の竹内綾乃氏を主たる参加メンバーとしている。研究を遂行するために、コアメンバー以外も自由に参加できる月例の研究会・討論会などを通して、研究情報の交換、論文や開発ツール、構想段階のプロジェクトや進行中のパイロットスタディのプレビューなどを行っている。

2022年度は、前年度に引き続き、10月までの研究会をオンラインで開催したが、11月に待望の対面開催(オンライン併用のハイブリッド方式)に移行し、以後はハイブリッド方式で開催した。オンラインでの研究会の開催を続けているうちに、学外からの研究会参加者が増加したこともあり、今後もハイブリッドでの開催を続ける予定である。

2022年度「テクストマイニングとデジタルヒューマニティーズ」研究会開催記録 およびメンバーによるDH関連学会での発表記録

第1回 2022年4月15日開催

発表者・発表題目

全メンバー 2022年度の活動計画打合せ

第2回 2022年5月20日開催

発表者・発表題目

田畠 智司 「Using topic models to shed new light on body language in Dickens's fiction」

第3回 2022年6月10日開催

発表者・発表題目

藤田 郁 「トピックモデルを用いた Alfred Tennyson の韻文研究」

第4回 2022年7月8日開催

発表者・発表題目

曹 芳慧 「Thomas Hardy, *Tess of the d'Urbervilles* の文体への語彙的アプローチ
徐 勤 「中国語作文における言語特徴の考察:
多次元分析による日本人中国語学習者と中国語母語話者の作文の比較」

第5回 2022年8月5日開催

発表者・発表題目

三野 貴志 「一般動詞を伴う there 構文
～コーパス調査から見えてくること～」
岩根 久 「Ronsard のソネの比較に向けて
Les Amours (1552) と *Les Amours* (1553)」

第6回 2022年9月2日開催

発表者・発表題目

八野 幸子 「教科等横断的視点を取り入れた英語教育支援ツール作成に向けた語彙研究」
山田 彰堯 「適用形の通時的構文交替:「させていただく」「させてもらう」「させてくださる」「させてくれる」の選択に対する状態空間モデルを用いた時系列分析」
Hodošček Bor 「英語電子書籍テキスト処理
'Standard Ebooks' と spaCy を例に」

第7回 2022年10月1日開催

発表者・発表題目

竹森ありさ 「色彩語 white を含む強意直喩表現の意味と媒体の背景について」
李 晨婕 「コーパスに基づくシャーロック・ホームズシリーズと
そのパスティーシュの文体的対比研究」

第8回 2022年10月2日開催 英語コーパス学会第47回大会
発表者・発表題目

黒田 純香 「トピックモデル可視化ツールの開発に向けて」
藤田 郁 “LDA Topic modelling of Tennyson's Poetry”（学生優秀発表賞受賞）

第9回 2022年11月4日開催
発表者・発表題目

菅原 裕輝 「研究紹介」
ハラルド ク Operationalizing topic models for writing conceptual history
マレ

第10回 2022年12月2日開催
発表者・発表題目

小堀 彩夏 「英訳版 村上春樹作品における特徴の比較」
福本 広光 「米国一般教書演説に見られる分離不定詞の実態と効果について」

第11回 2023年1月6日開催
発表者・発表題目

浅野 元子 「日本語論文の英語抄録における言語使用の調査：
日本語抄録の動詞的名詞に着目して」
王 錦 「日本語における2種類の分離動詞の違い
一分離元と分離物の関係および構文的意味についてー」
Vogatza Dim- 「イギリス文学におけるPITYの訳語の比較研究」
itra

第12回 2023年2月10日開催

発表者・発表題目

- | | |
|-------|--|
| 今尾 康裕 | 「CasualMallet の機能紹介」 |
| 竹内 紗乃 | 「登場人物のことばから探る人物造型論の可能性
—源氏物語における六条御息所—」 |

第13回 2023年3月15日開催

発表者・発表題目

- | | |
|-------|---|
| 涌井 萌子 | 「匿名政治パンフレットの計量的分析
—「レ枢機卿のマザリナード」の帰属検証—」 |
| 塚越 柚季 | 「サンスクリット文献『リグ・ヴェーダ』の韻律分析用データベースと
『リグ・ヴェーダ』の文書間類似度」 |
| 出川 英里 | 「KHCODER を利用した裁判判例の数量分析
—19世紀エジプトの判例集を素材として—」 |

2022年 2月
研究代表者 田畠 智司