The University of Osaka
Institutional Knowledge Archive

Title	「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
Author(s)	田畑, 智司
Citation	言語文化共同研究プロジェクト. 2022, 2021, p. 1-4
Version Type	VoR
URL	https://doi.org/10.18910/88356
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 彬尭講師の着任があり，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言語文化研究科の専任教員 5 名と名誉教授 1 名 (今尾 康裕, 田畑 智司, Hodošček Bor，三宅 真紀，山田 彬尭，岩根 久名誉教授)，当研究科博士後期課程在学生 7 名（黒田 絢香，黄 晨雯，岡部 未希，徐 勤，福本 広光，藤田 郁，竹森 ありさ)，文学研究科博士後期課程在学生 1 名（涌井 萌子)，博士前期課程在学生 3 名（王 鈺，曹 芳慧，李 晨婕)，研究生 2 名（Vogatza Dimitra, 李 建蕾）に加え，OG の大阪医科薬科大学 浅野 元子氏（2020 年 3 月博士学位取得）・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)，数理・データ科学教育研究センターの上阪 彩香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。

2021 年度は，前年度に引き続き，パンデミック下全ての研究会をオンラインで開催した。ほぼ 2 年間，対面開催がなかったためか，今ではむしろ，オンライン開催の方がデファクト・スタンダードになってしまった感もある。インフォーマルな雰囲気の中で情報交換を行う過程で思いがけない新たな着想，セレンディピティを生むオフライン，対面でのやり取りができなかったのは残念であったが，その一方で，研究資料やデータの共有が効率化されるとともに，オンラインでのプレゼンテーションのノウハウが蓄積され，学術活動の引き出しが増えるという好循環も生じたことは確かであろう。

2021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　2021年4月16日開催（Zoomを介したオンライン）
発表者・発表題目

全メンバー　2021年度の活動計画打合せ

第2回　2021年5月14日開催（Zoomを介したオンライン）
発表者・発表題目

涌井 萌子　「計量文体学を用いた匿名文書の帰属問題への挑戦
レ枢機卿のマザリナードを例に」

第3回　2021年6月11日開催（Zoomを介したオンライン）
発表者・発表題目

藤田 郁　“Topic modelling of Tennyson”
蒋　睿傑　「中国人日本語独習者がアニメに求める学習需要について
―人気アニメのセリフの特徴分析から―」

第4回　2021年7月2日開催（Zoomを介したオンライン）
発表者・発表題目

黄 晨雯　“What is popular in webnovels’ info?”
徐 勤　「中日大学生が書いた中国語作文における叙述視点
「外国語の勉強」と「留学」をライティングのテーマとして」

第5回　2021年7月7–9日開催（Zoomを介したオンライン）　Poetics and Linguistics Association International Conference (PALA) 2021, University of Notthingham, United Kingdom
発表者・発表題目

Tomoji Tabata　“Different paths to the same peak: Digital humanities and Spitzerian stylistics” (Keynote lecture)

第6回　2021年8月6日開催（Zoomを介したオンライン）
発表者・発表題目

浅野 元子　「ジャンルの理解度に関する考察
―サイエンスニュースを用いた学術英語への導入授業を例に―」
岩根 久　「Ronsardのソネの比較に向けて
Les Amours (1552) と *Les Amours* (1553)」

第7回　2021年9月3日開催（Zoomを介したオンライン）
発表者・発表題目

三野 貴志　「ThereAmalgam研究のこれから」
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」
田畑 智司　「ワークショップ：近代日本文学作品コーパスのテクスト分析」

第 8 回　2021 年 10 月 1 日開催（Zoom を介したオンライン）
発表者・発表題目

福本 広光　「米国一般教書演説に出現する分離不定詞の用法について
1960 年代から 2020 年までの分析」
王 鈺　「分断・破壊事象を表す動詞の使い分けとその習得
―「切る」と類義語の関連性について―」

第 9 回　2021 年 10 月 2 日開催（Zoom を介したオンライン）　英語コーパス学会第 47 回大会
発表者・発表題目

黒田 絢香　「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 10 回　2021 年 11 月 5 日開催（Zoom を介したオンライン）
発表者・発表題目

三宅 真紀　「シナイ写本の字形分析<ふたたび>
―異常検知からのアプローチ―」
岡部 未希　「Dickinson の詩における宝石の比喩
―詩中で鮮やかに輝く自然―」

第 11 回　2021 年 12 月 3 日開催（Zoom を介したオンライン）
発表者・発表題目

竹森ありさ　「強意直喩表現 as white (or pale) as a sheet の意味変遷」

第 12 回　2021 年 12 月 16–17 日開催（Zoom を介したオンライン）　Workshop of Education on Digital Humanities: Innovative Application in the Big Data Era (WEDHIA) 2021, National Chengchi University, Taiwan
発表者・発表題目

藤田 郁　““Were not Tennyson’s words delicious?”: An Analysis Using LDA Topic Modelling”
黄 晨雯　“Topic analysis for the synopsis of web novels
―Using the Top2Vec Topic Model―”
黒田 絢香　“Applying topic modeling to full-length novels:
effective settings and visualizations”

第 13 回　2022 年 1 月 7 日開催（Zoom を介したオンライン）
発表者・発表題目

黒田 絢香　「トピックモデル可視化ツールの開発に向けて」
曹 芳慧　「Thomas Hardy, *Tess of the d’Urbervilles* における語彙的特徴」

第 14 回　2022 年 2 月 4 日開催（Zoom を介したオンライン）
発表者・発表題目

Hodošček Bor　「英語電子書籍テキスト処理
‘Standard Ebooks’ と spaCy を例に
高橋 新　“Study on the Application of Stylometric Methods for Analysing English Translations of the Bible: 17 + 3 Gospels of MARK and JOHN”
山田 彬尭　「イ形容詞文における丁寧語使用の歴史的変化
状態空間モデルを用いた時系列分析」

第 15 回　2022 年 3 月 4 日開催（Zoom を介したオンライン）
発表者・発表題目

李 晨婕　（発表予定）
今尾 康裕　（発表予定）

2022 年　2 月
研究代表者　田畑 智司