OUKA
The University of Osaka
Institutional Knowledge Archive

Title	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. 2019, 2018, p. 1-4
Version Type	VoR
URL	https://doi.org/10.18910/72739
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 2018」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 年度に岩根教授，緒方助教授，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。2010 年にはサイバーメディアセンターの森助教が加わり，翌 2011 年には言語文化教育論講座に新たに着任した今尾講師が加入した。さらに，2014 年度後期から新メンバーとして Bor Hodošček 講師が加わり，現在の陣容ができあがった。(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，コロケーション，意味構造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（岩根　久，三宅 真紀，今尾 康裕，Bor Hodošček, 田畑 智司)，当研究科博士後期課程在学生 5 名（杉山 真央, 土村 成美, 浅野 元子 , 黒田 絢香, 三野 貴志)，博士前期課程在学生 3 名（廣瀬 由奈，福本 広光，岡部 未希）に加え，本学非常勤講師の高橋 新氏 , 南澤 佑樹氏（本研究科修了)，摂南大学後藤 一章氏（本研究科修了)，帝塚山学院大学八野 幸子氏（一昨年 3 月当研究科より博士学位授与)，本学データビリティフロンティア機構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，さらには統計数理研究所の言語系共同利用研究班との夏・春の合同セミナーの開催などを通して，研究情報の交換，論文や開発ツールのプレビューなどを行った。2018 年度の研究会の開催記録を以下に記す。

2018 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 2018」研究会開催記録

第 1 回　2018 年 4 月 27 日開催「2018 年度の共同研究計画についての打合せ」

第 2 回　2018 年 5 月 25 日開催
発表者・発表題目
上阪 彩香　「井原西鶴作品の計量文献学的研究の現在」
ホドシチェク ボル “Linguistic complexity in a multilingual setting: Using the STTR to quantify “brow”-ness.”

第 3 回　2018 年 6 月 15 日開催
発表者・発表題目
杉山 真央　“President under the Eyes of Russian Media: The Stylometry of Russian Presidential Addresses and their Media Coverage.”

第 4 回　2018 年 7 月 13 日開催
発表者・発表題目
廣瀬 由奈　「米国トランプ大統領の移民に対する談話ストラテジーの批判的談話研究
—墨国ペニャニエト大統領の移民に対する談話ストラテジーと比較して—」
福本 広光　“A Diachronic Study of English Split Infinitive: A Qualitative and Quantitative Analysis with Special Reference to Its “Splitters.””

第 5 回　2018 年 8 月 10 日開催
発表者・発表題目
岩根　久　「計量的視点からテキストを見る 同一ソネ内の共起単語ペア 」
ホドシチェク ボル “Reproducibility challenges in integration of Open Linked Data with the Aozora Bunko corpus”
浅野 元子　「英語学術論文執筆の負担に関する量的調査の予備検討：日本の理系大学院生および研究者を対象に」

第 6 回　2018 年 9 月 1–2 日開催（統計数理研究所共同利用研究班との合同中間報告会として開催
於 神戸大学百年記念館）
発表者
杉山 真央　「エリツィン大統領のロシア大統領年次教書演説と新聞報道」
浅野 元子　「国際英語としての日本人英語医学論文における言語的特徴」
高橋　新　「英語翻訳聖書間の計量的スタイル及び語彙選択の分析 (中間報告)—マルコ及びヨハネによる両福音書 17 翻訳の分析—」
田畑 智司　“Vector representation of words and text clusterings”
今尾 康裕　「学習者による語彙表現の過剰・過少使用分析における一考察」
土村 成美　「トピックモデルを用いた Agatha Christie 作品の特徴に関する分析」

八野 幸子　「教科横断的視点を取り入れた自己表現活動ための語彙研究」

第 7 回　2018 年 9 月 7 日開催

八野 幸子　「教科横断的視点を取り入れた，自己表現活動のための語彙分析―昆虫に関する語彙を中心に―」

南澤 祐樹　「概念メタファー理論から見た感情のプロトタイプ」

土村 成美　“A Quantitative Analysis of Agatha Christie’s Works Applying a Machine Learning Approach.”

第 8 回　2018 年 10 月 12 日開催

発表者・発表題目

土村 成美　「Agatha Christie と同時代作家との文体比較」(ポスター発表)

第 9 回　2018 年 11 月 9 日開催

発表者・発表題目

上阪 彩香　「著者判別分析における形態素解析辞書選択」

杉山 真央　“Russian Presidents versus Russian press: Applying text-mining approaches to explore different points of view”

第 10 回　2018 年 12 月 7 日開催

発表者・発表題目

田畑 智司　“Word vectors and semantic style”

第 11 回　2019 年 1 月 11 日開催

発表者・発表題目

福本 広光　“A Corpus-Based Analysis of Split Infinitives in American English: With Special Reference to “Splitters” from Diachronic and Rhetorical Viewpoints.”

第 12 回　2019 年 2 月 8 日開催

発表者・発表題目

今尾 康裕　「英日バイリンガルエッセイコーパスで遊んでみる」

第 13 回　2019 年 3 月 8 日開催

発表者・発表題目

三宅 真紀　「シナイ写本写字生 A/D の識別ーエプシロンおよびニューを対象としてー」

高橋　新　“Study on the Application of Stylometric Methods to Analysing English Translations of

the Bible (An Interim Report): Gospels of Mark and John.”

岡部 未希　“*Thou* & *You* in Emily Dickinson’s Poems: Focusing on Interjections.”

第 14 回　2019 年 3 月 20–21 日「言語研究と統計 2019」（於 統計数理研究所）として開催

発表者・発表題目

浅野 元子　「日本人著者による英語医学論文のムーブの検討―Nwogu の研究を踏まえて」

今尾 康裕　「名詞修飾に関する一考察」

上阪 彩香　「西鶴浮世草子の文章の特徴と出版時期」

岡部 未希　*Thou* & *You* in Emily Dickinson Poems: Focusing on interjections

黒田 絢香　「トピックモデルによる特徴語抽出の試み」

後藤 一章　「テキストジャンルにおける名詞用法の差異」

高橋　新　「英語翻訳聖書間の計量的スタイル及び語彙選択の分析 ― マルコ及びヨハネによる両福音書 17 翻訳の分析 ―」

田畑 智司　「Word Vectors and Semantic Style in Classic Fiction」

土村 成美　「Agatha Christie 作品の分析」

八野 幸子　「教科横断的視点を取り入れた英語教育のための語彙研究」

福本 広光　「アメリカ英語における分離不定詞の通時的使用実態の研究」

南澤 佑樹　「名詞 emotion のコロケーション：概念メタファー理論の観点から」

三宅 真紀　大阪大学「シナイ写本の写字識別の試み：大文字エプシロンおよびニューの字体に注目して」

第 15 回　2019 年 3 月 27 日 Corpus Stylistics Forum with Dan McIntyre and Emerging Scholars として開催

発表者・発表題目

Motoko Asano,“Publishing Research in English: A Replication Study of Nwogu’s 1997 Work with a Specialized Corpus of Japanese-Authored Medical Articles”

Miki Okabe,“*Thou* & *You* in Emily Dickinson’s Poems: Focusing on interjections”

Narumi Tsuchimura, “Exploring Stylistic Features in Agatha Christie’s Works: Comparing with Her Contemporaries”

Ayaka Kuroda, “Machine Learning Analysis of Detective Fiction: Using Topic Models to Extract Key-Words”

Yuki Minamisawa, “Centrality of Metaphors and Metonymies: The Case of ANGER”

Special lecture: Dan McIntyre (University of Huddersfield, UK), “Just what is corpus stylistics?”

2019 年 5 月

研究代表者　田畑 智司