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. 2017, 2016, p. 1-4
Version Type	VoR
URL	https://doi.org/10.18910/62035
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, 南澤 佑樹, 夏　天驕, 土村 成美, 浅野 元子)，博士前期課程在学生 2 名（黒田 絢香, 藤原 いずみ）に加え，本学非常勤講師の高橋 新氏，摂南大学の後藤 一章氏（本研究科修了)，和歌山大学非常勤講師の木山 直毅氏（2016 年度当研究科より博士学位授与)，帝塚山学院大学の八野 幸子氏（博士論文執筆中)，同志社大学の上阪 彩香氏を主たる構成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。2016 年度の研究会の開催記録を以下に記す。

2016 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

第 1 回　2016 年 5 月 13 日開催「2016 年度の共同研究計画についての打合せ」

第 2 回　2016 年 6 月 10 日開催

発表者・発表題目

夏　天驕　「中国語の事象アスペクトと構文：MDS による分析」

杉山 真央　「Diachronic changes of the Russian Presidential Addresses to the Federal Assembly: Comparing a perspective of archetype key words」

第 3 回　2016 年 7 月 1 日開催

発表者・発表題目

南澤 佑樹　「怒りの感情を表すメトニミー的表現」

杉山 真央　（ポスター発表）“Diachronic changes of the Russian Presidential Addresses to the Federal Assembly: Comparing a perspective of archetype key words”

第 4 回　2016 年 8 月 5 日開催（来阪中の Auckland 大学 Michael Barlow 先生を招いて）

発表者・発表題目

岩根　久　「語彙計量的手法を日常のテキスト分析に」

土村 成美　“Stylistic Analysis of Agatha Christie’s Works: Comparing with Dorothy Sayers”

浅野 元子　“Linguistic Features of English Medical Research Paper Genres and Their Pedagogical Implications: A Pilot Study”

第 5 回　2016 年 8 月 29–30 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）

発表者

浅野 元子　「医学論文の言語的特徴についてのコーパス研究 : ESP と国際英語としての探索」

八野 幸子　「自己表現活動のための語彙に関する研究」

今尾 康裕　「bi-gram 分析の可能性を探ってみる」

高橋　新　「英語翻訳聖書間の計量的スタイル分析手法の考察と今後の研究視点」

上阪 彩香　「浮世草子作品の電子データ化について」

木山 直毅　「英語の小説などに見られる引用句倒置の観察」

Bor Hodošček「日本語作文推敲支援システムのためのウエブベース作文環境の検証」

三宅 真紀　「新約聖書写本間における異読距離計算の進捗状況」

岩根　久　「反ロンサールパンフレットの計量的特徴を探る」

田畑 智司　「Body-part words を通して見る fiction の言語」

第 6 回　2016 年 9 月 6 日開催

高橋　新　“A Study on the Application of Stylometric Methods for Analysing English Translations of Bible”

八野 幸子　「Corpus of Global Web-based English Full text 版の分析による自己表現活動のための語彙研究の試み」

木山 直毅　「英語小説における引用句倒置」

第 7 回　2016 年 10 月 7 日開催

発表者・発表題目

Bor Hodošček「日本語作文推敲支援システムのためのウエブベース作文環境」

第 8 回　2016 年 11 月 11 日開催

発表者・発表題目

弘田 愛咲子“A Study on Quotative Inversion”

第 9 回　2016 年 12 月 9 日開催

発表者・発表題目

田畑 智司　（ポスター発表 1）“Sequencing a Literary Genome of Classic Fiction: When the Humanities Meets Digital”

（ポスター発表 2）“Collaborative Texts under a Stylometric Microscope: Investigating Cases of Mixed Authorship”

第 10 回　2017 年 1 月 20 日開催

発表者・発表題目

後藤 一章　「機械翻訳を利用した英日対訳表現抽出の試み」

木山 直毅　「引用表現の語順と動詞のカテゴリー」

第 11 回　2017 年 2 月 10 日開催

発表者・発表題目

今尾 康裕　「次なるステップへ—依存文法を利用したコロケーション検索の可能性」

上阪 彩香　「西鶴および団水の比較分析」

第 12 回　2017 年 3 月 10 日開催

発表者・発表題目

三宅 真紀　「人工写本 Parzival を利用した校合データ様式と異読距離の考察」

黒田 絢香　「小説テクストの計量的分析：アーサー・コナン・ドイルの作品から」

第 13 回　2017 年 3 月 27–28 日「言語研究と統計 2017」（於 統計数理研究所）として開催

発表者・発表題目

浅野 元子　「医学論文考察部に学ぶ：論理展開の出現と特徴語についての検討」

木山 直毅　「英語の引用句構文に見られる動詞の特徴」

土村 成美　「Agatha Christie 作品の統計的文体分析」

岩根　久　「反ロンサールパンフレットの計量的特徴の再検討」

上阪 彩香　「アンサンブル学習モデルを用いた近世文学作品の著者に関する検討」

三宅 真紀　「人工写本 Parzival を利用した校合データ様式と異読距離の考察：新約聖書写本比較に向けて」

高橋　新　「英語翻訳聖書間の計量的スタイル分析手法の考察―「ヨハネによる福音書」の分析を通して―」

藤原 いずみ「宮沢賢治の経年的な文体の変化―計量的な分析を用いて―」

黒田 絢香　「小説テクストの計量的分析―アーサー・コナン・ドイルの作品から」

今尾 康裕　「構文解析を利用した英語コロケーション分析の可能性」

田畑 智司　“The Semantic Universe of Classic Fiction”

2017 年 4 月

研究代表者　田畑 智司