

二国間交流事業 共同研究報告書

平成 23 年 10 月 28 日

独立行政法人日本学術振興会理事長 殿

共同研究代表者所属・部局 大阪大学・大学院言語文化研究科

職・氏名 (ふりがな) 准教授 田畑 智司
たばた ともじ

1. 事業名 相手国 (英国) との共同研究 振興会対応機関 (British Academy)

2. 研究課題名 多変量文体分析モデルによる 19 世紀英国新聞の計量分析研究

3. 全採用期間

平成 21 年 4 月 1 日 ~ 平成 23 年 9 月 30 日 (2 年 6 ヶ月)

4. 経費総額

(1) 本事業により執行した研究経費総額 2,000,000 円

初年度経費 1,000,000 円、 2 年度経費 1,000,000 円、 3 年度経費 円

(2) 本事業経費以外の国内における研究経費総額 0 円

5. 研究組織

(1) 日本側参加者（代表者は除く）

氏名 <small>(ふりがな)</small>	所属・職名	研究協力テーマ
みやけ まき 三宅 真紀	大阪大学・大学院言語文化研究科・准教授	ネットワークモデルによるテキスト種類の視覚化に関する研究
こばやし ゆういちろう 小林 雄一郎	大阪大学・大学院言語文化研究科・言語文化専攻博士後期課程 3 年次在籍	テキストマイニングを応用したテキストの言語研究

(2) 相手国側研究代表者

所属・職名・氏名 英国 King's College London Centre for Computing in the Humanities・Professor・Harold Short

(3) 相手国参加者（代表者は除く）

氏名	所属・職名（国名）	研究協力テーマ
Gerhard Brey	King's College London Centre for Computing in the Humanities・上級研究員 (英国)	人文学資料のデジタル化に関する研究
John Lavagnino	King's College London Centre for Computing in the Humanities・Reader (英国)	Digital Culture/Digital Humanities
José Miguel Monteiro Vieira	King's College London Centre for Computing in the Humanities・Research Associate (英国)	電子出版フレームワーク, 自然言語処理, 視覚化方法論
Matteo Romanello	King's College London Centre for Computing in the Humanities・Ph. D candidate (英国)	自然言語処理技術を応用した知識獲得に関する研究

6. 研究実績概要（全期間を通じた研究の目的・研究計画の実施状況・成果等の概要を簡潔に記載してください。）

本研究は、大阪大学大学院言語文化研究科のデジタルテキスト分析研究チーム(3名)と、英国 King's College London の Centre for Computing in the Humanities (コンピュータ利用人文科学センター、以下 CCH と表記)に在籍する研究者(5名)による、19世紀の英国新聞データベース(NCSE)の計量分析に関する共同研究である。デジタルテキストの情報付与法に関する国際共通規格 Text Encoding Initiative の開発、応用研究において学界の中核的な役割を担う CCH が編纂した大規模な言語データベースや電子テキストの情報付与法に関する豊富な知見と、大阪大学において申請者および研究分担者が確立し、精緻化を進めているテキストの計量分析方法論、特に、多変量文体分析モデルおよびラフ理論に基づくネットワーク分析モデルに関する知見とを有機的に組み合わせることにより、英国 19 世紀の新聞の言語特徴やキーワードの使用動態を定量的に調査することを目的とした。

本研究では、多変量文体分析モデルおよびネットワーク分析モデルを応用したテキスト分析を実施した。研究の過程で NCSE にはこれまであまり把握されていなかった問題が数多く判明した。特に、古いジャーナル誌面の印刷品質に起因する OCR エラーを数多く発見し、CCH 側にフィードバックし、今後の NCSE の改訂について課題を明確化することに貢献した。こうした点を踏まえ、2年目の研究は NCSE のうち *English Woman's Journal* に焦点を絞って分析を行った。その結果、(1) 重要語彙項目と共起する多数の語彙項目間の相互関係、(2) 重要語彙項目が生起するテキスト間、ジャンル間の相互関係、さらに、(3) 重要語彙項目、共起語とテキスト、ジャンルとの相互関係を視覚化することに成功した。特に、米国で初の女性編集者による刊行誌 *American Ladies' Magazine* と比較すると、*English Woman's Journal* には、女性の話し言葉の言語特徴がより色濃く反映していることなどを明らかにした。このようにして、*English Woman's Journal* に通底する特徴や、その言語文化学的特徴を俯瞰的に分析した。

本研究の成果は、大学共同利用法人・情報システム研究機構・統計数理研究所において平成 21 年から 23 年にかけて行われた 3 回の研究セミナー「言語研究と統計」で発表したほか、本研究プロジェクトを核にした国際シンポジウム *Osaka Symposium on Digital Humanities 2011* を大阪大学にて開催した。当シンポジウムでは、*Statistical text-mining on English Woman's Journal* と題するパネルセッションを企画し、研究代表者および研究協力者それぞれが成果発表を行った。