

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 13 日現在

機関番号：17401

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26580053

研究課題名(和文) A Digital Humanities Analysis of Christianity in the 19th-Century British Fiction

研究課題名(英文) A Digital Humanities Analysis of Christianity in the 19th-Century British Fiction

## 研究代表者

大野 龍浩(OHNO, Tatsuhiro)

熊本大学・文学部・教授

研究者番号：80169028

交付決定額(研究期間全体)：(直接経費) 1,000,000円

研究成果の概要(和文)：19世紀英国の著名な28作家による291編の小説をreference corpus、主要9作家ごとの小説群をstudy corpusにして、effect sizeを計算し、上位1.0%の単語を抽出後、15の範疇に分類したうち、「キリスト教道徳」に属する単語に着目して、effect sizeの高い順に列挙すると、G. Eliot, Gaskell, Austen, Gissing, C. Bronte, Hardy, A. Bronteの順になった。DickensとE. Bronteは、reference corpusとの有意な差はなかった。作家ごとのキリスト性的一端を示す新データである。

研究成果の概要(英文)：The top-ranking keywords obtained through the effect-size analyses of the 9 major novelists' study corpora in comparison with the reference corpus of the 291 novels by the 28 eminent 19th-century British authors were classified into 15 categories according to the connotation of each word. The effect size of the author's words in the Christian-morality group was higher in the order of G. Eliot, Gaskell, Austen, Gissing, C. Bronte, Hardy, and A. Bronte. Dickens and E. Bronte showed little difference from the reference corpus in their use of Christian-morality words. This result sheds light on the individual novelist's unique description of Christian morality which has "for a long time . . . received only impressionistic and subjective treatment" (David I. Homes). Its uniqueness should be confirmed could this outcome be validated by further close reading of their texts.

研究分野：英文学

キーワード：Digital Humanities Effect Size Topic Modelling Corpus Stylistics Christianity

### 1. 研究開始当初の背景

「文学作品の解釈には正解はなく、作品は読者の自由に解釈でき、優れた作品は読みの多様性を許容する」というのが、文学批評の常識とされる。そうして、批評家は作品を自らの関心で切り取り、その切り取った対象なり、作品の背後にある歴史的・社会的背景なりを明らかにする。

本研究はこの文学批評の常道とは違った方向から文学作品の研究を試みることから出発した。つまり、唯一真の解釈すなわち、原作者が意図した解釈、「絶対解釈」(the absolute interpretation)の可能性を探ることを目標にした。

原作者の意図を重視する解釈学者(intentionalist hermeneuticists)によれば、文学 text には2面性があるという。時代や地域によってその持つ意義が変わる側面と、変わらない側面。たとえば、Jane Austen の *Pride and Prejudice* (1813) であれば、19世紀初頭の英国民に与える影響と、当時の歴史、社会、文化的背景を知らない21世紀初頭に住む日本人に与えるそれとは違う。同時に、作品が与える感動には時代や地域を超えて、人類に共通するものがある。Intentionalist hermeneuticists たちは、前者を作品の「意義」(significance) 後者を作品の「意味」(meaning)と呼んで区別する。そして、それぞれを探求する作業を批評(criticism)と解釈(interpretation)と定義する。また、いわゆる author にも2種類あるとし、それは作品を実際に書いた人間と作品を精読することによってあぶり出されてくる作者で、前者を「作者」(author) 後者を「原作者」(author construct)と呼ぶ。探るべき作者の意図は、後者の意図で、前者の意図ではない。何故なら、作品は必ずしも作者が予定していたプロット通りに進むとは限らないし、完成した作品にこそ原作者の意図が内包されているからである。従って、彼らによれば、絶対解釈を探るためにすべきことは、原作者が作品に込めた「意味」を「解釈」することになる。伝統的な作品分析法は、作品の「意義」を「批評」することとなる。

この目的を達成するために最も確実な方法のひとつは、単語を始めとする作品構造を客観的に明らかにすることである。なぜなら、作者は主題を最も効果的に読者に伝えるために、単語を選び、登場人物、作品の舞台となる時間や場所などの作品構造を決定するのだから。

こういうわけで、方法論としては、コーパス文体論(corpus stylistics)という、デジタル人文学(Digital Humanities)の一手法を用いる。すなわち、対象作品の e-texts をテキスト分析ツール(AntConc)に読み込ませて得られる、作品に用いられたすべての単語に関するデータを基に、類似項目ごとに分類するなどして、その使用傾向の全体的な特長を明らかにする。

こうすることによって、原作者ごとの作品構成上の違いを明らかにしたり、原作者の意図をできるかぎり客観的に類推することに挑戦する。

### 2. 研究の目的

19世紀英国小説にキリスト教がどう描かれているかを、Digital Humanitiesの手法(distant reading)を用いて客観分析したあと、その妥当性をclose readingによって吟味するのが本研究の主たる目的である。

### 3. 研究の方法

主として用いた方法は、effect size と topic modelling という Digital Humanities でよく用いられる text mining の方法である。

Effect size 分析で対象にしたのは、著名な28作家による291編の小説。すべての e-texts を集めたものを reference corpus にし、作家ごとの小説群を study corpus にして、effect size を計算し、上位1.0%を抽出。その後、各々の単語が持つ意味合い(connotation)から、15グループ(GOD\_morality, GOD\_nature, HUMAN BEINGS\_action, HB\_characters, HB\_clothing, HB\_condition, HB\_household, HB\_people, LANGUAGE\_dialect, LANGUAGE\_French, LANGUAGE\_others, PLACE\_buildings, PLACE\_others, PLACE\_place names, TIME)に分類。Effect size は study corpus 中のある単語の数と reference corpus 中のそれとの差を100万語あたりに換算して示す数値だから、その値が高いほど、その単語はその作家特有のものということになる。

Topic modelling 分析で対象にしたのは、Elizabeth Gaskell の短編小説。MALLET というフリーのソフトウェアに全52編の e-texts を読み込ませ、上位に来る topic を特定する。MALLET には運用するたびに若干違った結果を提示する特徴があるので、信頼性を確保するため、10回操作して得られたデータについて、最も高い配分率の topic の最も高い配分率の単語を特定した。Topic とは特定の文書内で同時に使われる頻度の高い単語の集合体の中で、MALLET は使われた単語の頻度を統計的に処理しながら、関連性のある単語の集まり(topic)を推定していく。同一文書の中で現れる topic が多いほど、その topic の同一文書内での配分率は高くなる。

### 4. 研究成果

上記のうち、主要9作家(Jane Austen, Elizabeth Gaskell, Charles Dickens, Charlotte Brontë, Emily Brontë, George Eliot, Anne Brontë, Thomas Hardy, and George Gissing)の小説に現れる effect size 上位の単語を15項目に分類した結果は、Fig. 1の通り。

これによると、19世紀小説家の標準と比べて、特有の単語を用いているのは、Hardy, Gissing, G. Eliot の順に高いことがわかる。

Hardy は場所に関する単語が多く、Gissing は人間の行動に関するそれがきわめて多い。G. Eliot も人間の行動に関する単語を Gissing に次いで多く使用している。それに、Austen と Gaskell が続く。Dickens は人物に関する単語をより多く使っている。

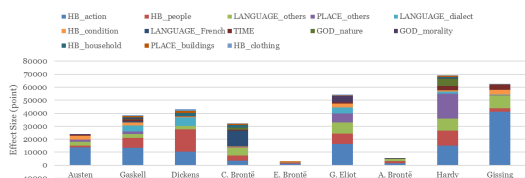


Fig. 1. Effect Sizes of the 9 Novelists' Keywords in 15 Categories

このうち、「キリスト教道徳」の範疇 (GOD\_morality) に分類された単語に着目して、主要9作家を effect size の高い順に列挙すると、G. Eliot, Gaskell, Austen, Gissing, C. Brontë, Hardy, A. Brontë の順になった。Dickens と E. Brontë は、reference corpus との有意味な差はなかった (Fig. 2 参照)。

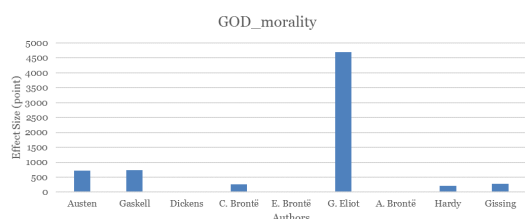


Fig. 2. Effect Sizes of the 9 Novelists' Keywords in "God\_Morality"

具体的な単語は、G. Eliot の場合、frate, preaching, rector, preacher, vicar, christian, spiritual, divine, tenderness, religion, souls, pity, religious で、牧師、宗教、靈魂、道徳など多岐のジャンルにわたる。Gaskell の場合、bless, comfort, sin, tender, thankful の5語で、人間の生き様に関する単語が最も使用頻度の高いキリスト教道徳関連の単語だった。Austen の場合は、affection, kindness, comfort の3語で、ともに人間関係に関わる単語。C. Brontë の場合は、spirit, affection, pure の3語で、人間精神に関係する単語。A. Brontë は God, Hardy は vicar, Gissing は sincerity のそれぞれ1語ずつ。篤い信仰の持ち主だった A. Brontë、牧師を折りに触れて作品に登場させた Hardy、誠実な人間の苦悩を描いた Gissing と、それぞれの特長がこのリストに凝縮されている。

Fig. 3 は、MALLET を10回操作して上位に来た topics と単語のデータを、x軸に全文書中の topic 配分率、y軸にその単語を含む文書の数をとって視覚化したものである。これによると、ギaskell短編の主要な topic には、time, day, house, father, mother などの単語が高頻度で含まれていることがわかる。こ

のことは、ギaskellの短編が、時間の経過 (time, day, long, life, night) のなかで繰り返される家族のコミュニケーション (house, man, good, father, thought, heard, looked, face, home, eyes, told, mother, knew) を描写する傾向にあることを示唆している。

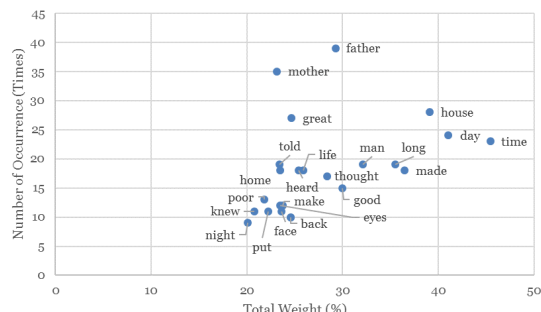


Fig. 3. Top 24 Words by Total Weight and Number of Occurrence in 10 Topic-Modelling Operations for Gaskell's 52 Shorter Fiction

同じデータのうち、god という単語を含む6つの topics に着目して、同様の処理をし、総配分率と文書数の相関性を調べたのが Fig. 4 である。この結果から、god を含むトピックは、child, dead, heart, words, father, eyes, life などの単語を同時に含む傾向が高いことがわかる。つまり、god への言及は、家族の絆 (child, words, father, life, mother, love, speak, knew, heard, spoke, woman) 肉体の部位 (heart, eyes, arms, face) 死 (dead, death) に関連して、なされる蓋然性が高いことが推察できる。この統計分析の結果をテキストと照合することによって、質的な裏付けを示した。

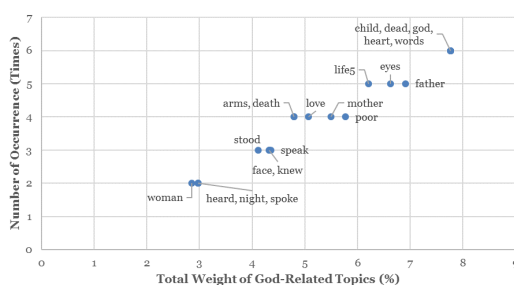


Fig.4. Top 21 Words by Total Weight and Number of Occurrence in 10 Topic-Modelling Operations for Gaskell's 52 Shorter Fiction

テキスト分析に Digital Humanities の手法を用いることには、テキストが結局は作者が選択した単語の集合体である以上、それに込められた構造上のパターンを比較的客観的に分析できるメリットがある。分析結果 (量的発見) は、テキスト内の文脈と照合すること (質的分析) によって相互補完させると、

信頼性が高まる。

今回の研究結果の意味するところをより正確に測るため、effect-size 分析で明らかになった使用頻度の高い単語の実際の使用例を topic modelling で確認したり、さらに綿密な close reading によってより詳細に検証したりすることが、今後の課題である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計0件)

[学会発表](計4件)

大野龍浩、「ブロンテ文学の科学的分析」(ブロンテ・デイ公開講座、横浜市立大学、2015年6月7日)

Tatsuhiko OHNO, “An Effect-Size Analysis of Christianity in the 19th-Century British Novels” (JADH2015、京都大学、2015年9月2日)

Tatsuhiko OHNO, “Digital Humanities Analysis of the 19th-Century British Novels” (Research Seminar, University of Birmingham, UK, 24 Sept 2015)

大野龍浩、「シンポジウム：ギャスケル中・短編小説の魅力—『没後150年記念論文集』を語る」(日本ギャスケル協会第27回大会、名古屋大学、2015年10月3日)

[図書](計4件)

“A Topic-Modelling Analysis of the Sacred and the Secular in The Life of Charlotte Brontë.” *Evil and Its Variations in the Works of Gaskell: A Sesquicentennial Commemorative Volume*. Ed. Mitsuharu Matsuoka. Osaka: Osaka Kyoiku Toshō, 2015. 209-24.

“An Effect Size Analysis of Christianity in the Fiction of the Brontë Sisters.” 『ブロンテと19世紀イギリス』日本ブロンテ協会編 大阪：大阪教育図書、2015. 75-98.

「コーパス文体論によるギャスケル短編作品の解析：そのキリスト教性について」『エリザベス・ギャスケル中・短編小説研究』日本ギャスケル協会編 大阪：大阪教育図書、2015. 41-52.

「ブロンテ文学の科学的分析」『文芸礼賛：アイデアとロゴス』大阪：大阪教育図書、2016. 479-95.

[産業財産権]

出願状況(計0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

[その他]  
ホームページ等

#### 6. 研究組織

(1)研究代表者

大野 龍浩 (OHNO, Tatsuhiko)  
熊本大学・文学部・教授  
研究者番号：80169028

(2)研究分担者

なし

(3)連携研究者

なし