

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年6月13日現在

機関番号：32631

研究種目：基盤研究(C)

研究期間：2008～2012

課題番号：20520240

研究課題名（和文） デジタル情報検索を活用した『フィネガンズ・ウェイク』の  
テキスト生成過程研究研究課題名（英文） Online Newspaper Archives and Google Book Search as *Finnegans Wake*  
Genetic Research Tools

## 研究代表者

扶瀬 幹生 (FUSE MIKIO)

聖心女子大学・文学部・教授

研究者番号：60262094

研究成果の概要（和文）：インターネットで購読できる歴史的新聞の電子アーカイブと Google Book Search の古書籍全文テキスト検索のサービスをフルに活用することにより、James Joyce, *Finnegans Wake* のテキスト生成過程研究の基幹となるノートブックの出典研究を画期的に進展させた。また出典研究のための必要基盤の一つとして 2006 年発見の Jahnke Notesheet 計 56 葉の判読・電子データ化と、保管時に乱れてしまっているページ順序の復元を達成した。さらに将来この研究を共同研究の形で一層効率化・加速化することを期して、*Finnegans Wake* のテキスト生成過程研究用電子アーカイブの試作に着手した。

研究成果の概要（英文）：The project has proved how Online Newspaper Archives and Google Book Search are both efficient tools for discovering the source newspaper articles and source books/articles of James Joyce's Buffalo Notebooks. As byproducts the project has also achieved the digitization of Joyce's Jahnke Notesheets and the establishment of a prototype of a prospective *Finnegans Wake* genetic research archive.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	500,000	150,000	650,000
2009年度	200,000	60,000	260,000
2010年度	200,000	60,000	260,000
2011年度	200,000	60,000	260,000
2012年度	200,000	60,000	260,000
総計	1,300,000	390,000	1,690,000

研究分野：人文学

科研費の分科・細目：文学、英米・英語圏文学

キーワード：国際情報交換(西ヨーロッパ・北米)、英米文学、アイルランド文学、ジョイス、電子メディア、電子アーカイブ、genetic criticism

1. 研究開始当初の背景  
研究代表者は James Joyce, *Finnegans Wake* のテキスト生成過程研究を十数年来継続し

て行ってきた。当初は書籍としての *James Joyce Archive* 所収のファクシミリ資料を判読のうえ電子化してエクセルに蓄積し、これ

にメタファイル検索をかけて諸データの関係性を分析する作業を主に行っていたが、インターネットに書籍の電子テキストが大量に出るようになった結果、特に Buffalo Notebook (Joyce がマニュスクリプトで使用する素材として新聞・書籍から転記した記事を記述した作業ノート。以下ノートブックと呼ぶ) の出典研究において Google Web Search を使ったオンライン検索の有用性にいち早く注目し、これを活用して一定の成果をあげていた。

## 2. 研究の目的

今回の研究では近年インターネットで閲覧・検索が容易にできるようになった歴史的新聞の電子アーカイブと古書籍の全文テキスト検索のサービス (Google Book Search) をフルに活用することによって、従来の研究水準をはるかに上回る精度と網羅性を伴ったかたちでノートブックの出典研究を進展させることを企図した。

さらに、当初研究計画には入れていなかったが、研究年度内に新たに得た機会を捉えて、次の二つを研究課題に追加した：

### (1) 研究年度4年目にノートブック

VI. C. 3 の研究に着手したが、このノートブックに記載されている事項の厳密な分析には、そこから相当数の項目が転記されている Jahnke Notesheet という 2006 年発見の 56 葉の自筆ノートシートのデータを参照することが必須であることが分かった。幸い研究協力者 Geert Lernout の仲介によって、Zurich James Joyce Foundation からこのノートシートのファクシミリコピーを入手でき、またこのノートシートの記載を研究中の米国の大学院生 Nicholas Morris から判読データの提供を受けることができたので、VI. C. 3 を含む C シリーズノートブック研究に不可欠な研究基盤として Jahnke Notesheet の判読と電子データ化を行った。

(2) また、研究最終年度に国立情報学研究所の外来研究員として電子アーカイブ制作の技術を学ぶ機会を得たので、当研究を将来オンライン上に用意した *Finnegans Wake* のテキスト生成過程研究用アーカイブを使って共同作業の形でさらに効率的に推進するための基盤づくりにも着手した。

## 3. 研究の方法

毎年一ないし二冊のノートブックを判読し、歴史的新聞の電子アーカイブと Google Book Search を活用して出典発見の作業をすすめ、さらに出典の発見によって判読の精度が向上するノートブックの項目がその後マニュスクリプトのどの箇所で使用されているかを分析判定する作業を進める形で、エクセルのデータを補充・改訂していった。

## 4. 研究成果

五年間の研究の結果、ノートブックの出典研究を画期的に進展させるという所期の目的は達成できた。以下「新聞アーカイブ検索」「Google Book Search」「電子アーカイブの試作」の三項目それぞれについて研究成果と今後の課題を記す。

### (1) 新聞アーカイブ検索

本研究で活用した新聞アーカイブは *Irish Times Digital Archive*、*Irish Newspapers Archive*、および *Times Newspaper Archive* の三つである。これらを活用した結果、*Irish Times*、*Irish Independent*、*Freeman's Journal*、*Connacht Tribune*、*Times* の各紙の記事を出典とするノートブックの項目を多数発見した。

その中には従来図書館で所蔵の現物の紙面やマイクロフィルムを使った研究で発見されていた出典記事と同一ページや同一の日付の紙面に新たに見つかった (すなわち従来の肉眼での走査による方法が見落とししていた) 出典記事も多く含まれ、従来の形の研究をリードしてきた Vincent Deane (本研究の研究協力者の一人) から、デジタルメディアの新聞アーカイブを利用した新しい研究方法の有効性についての認証が得られた。

一般に日付の付いた新聞記事をノートブックの出典として特定することに成功すると、そのノートブックの作成年代を推定する重要な手掛かりが得られたことになるが、特にこれまでこの手がかりが全く得られていなかった D シリーズのノートブック (ジョイスの自筆のノートブックとしては散逸して現存せず、それをラファエル夫人が部分的に転写した C シリーズのノートブックに間接的に痕跡をとどめているもの) について新聞記事の出典を発見できたのは、*Finnegans Wake* のテキスト生成過程研究史上初めてのことである。

このように新聞アーカイブを活用することで、これまでになく精度と効率性を伴った画期的な出典発見の道が開けることが実証できた一方、実際に新聞アーカイブを利用してみるといくつもの問題点も明らかになった。

新聞記事が OCR によって電子化されることで検索や閲覧が容易になるのは書籍の場合と基本的に同じだが、新聞記事の場合印字の状態が全体に不鮮明・不統一であったり、レイアウトが複雑であったり、広告のロゴのように整然としたテキスト情報としては簡単にとりだせない情報も多数含まれる。使用した新聞アーカイブはいずれも機械による文字認識を行っただけで、その読み取り結果に多数含まれる誤認識を人間の目によって修

正することまではしていない。この点で、新聞アーカイブの検索は真に網羅的とは言えない。この不足を補うために *Irish Newspapers Archive* ではワイルドカード検索やファジー検索の機能も用意されているが、その実用度は期待されるほど高くはない。また *Irish Times Digital Archive* では一定期間(月単位)の記事について、読み取り自体が全面的に失敗している、すなわち検索の対象にならないページが少なからずあることも判明した(前述の Vincent Deane の指摘による)。この不備を補うためには、結局研究者が記事の一つ一つ肉眼で読んでいく作業に依然として多くの時間を費やさなければならないことがわかった。

出典として発見した記事をどのような形で手元にデータとして保存・蓄積していくかについても課題を残した。今回は該当する記事の部分だけを画像としてキャプチャし、それをマークアップ可能な pdf ファイルの形で保存していく方法をとったが、将来さらに見落としや新たな新聞記事の出典発見を緻密に行うためには、少なくとも出典の記事が存在する紙面については全ページを手元に保存・蓄積していきたいし、さらに将来的には、研究用電子アーカイブのなかで該当する日付の紙面全体を検索・閲覧できるようにするのが理想だが、紙面を一ページ単位でまると OCR にかける高解像度の画像としてダウンロードできるサービスほどの新聞アーカイブも提供しておらず、またそもそも研究用電子アーカイブに複製を作ることは著作権上の問題もからむので実現は難しい。

以上の不備・不足点については今後の新聞アーカイブの技術面・サービス面の進展を注視して、問題解決の糸口をつかみたい。

## (2) Google Book Search

ノートブック項目にある語句の出典となる書籍を Google の Web Search を使って発見するという研究方法をすでに確立していたことは始めに記したとおりだが、その後 Google Book Search のサービスが年々充実して、そこで検索できる古書籍の量が飛躍的に増大し、Google Book Search によって出典を発見する頻度が高まってきていたので、今回の研究ではこのサービスによって見つかる出典ないし出典であることが疑われる書籍を実際にオンライン書店を通して購入したうえで OCR にかける、マークアップ可能な pdf ファイルとして保存したものを使ってその書籍に存在する出典箇所を遺漏なく洗い出すという研究プロセスを確立した。その結果、実際に出典となる書籍を多数発見することができた。本研究の2年目から研究協力者に加わった Robbert-Jan Henkes がこの研究方法に習熟し、毎年重要な書籍出典を発見するこ

とで本研究に多大な貢献をしたことも明記しておきたい。

発見した書籍の出典として特記すべきものは以下の通り:

① 研究代表者がノートブック VI. B. 3 および VI. B. 14 にみられる Oscar Wilde に関する項目の出典が雑誌 *Studies: An Irish Quarterly Review* 13. 49 (March 1924) 所収の Herbert Thurston, “The ‘Oscar Wilde’ Script in Its Bearing on Survival.” であることを突き止め、出典が Hester Travers Smith, *Psychic Messages from Oscar Wilde* (1924) であるとする従来の定説を覆した。これに加えて、同雑誌からその他多数の出典論文を発見し、この二冊のノートブックの判読不明箇所の解明が大いに進んだ。

② 研究協力者の Robbert-Jan Henkes がノートブック VI. D. 1 の主要な出典として 1910 年代以降刊行された *To-Day and To-Morrow* 叢書の九冊を発見した。この発見は前述の研究代表者による新聞記事出典の発見とあわせて、散逸した D シリーズのノートブック研究に重要な貢献をした。

なお出典書籍の全ページを手元に保存・蓄積する方法については、前記の新聞記事出典の場合と違って特に困難はない。ただし、将来的には後述の電子アーカイブの中に組み込まれた XML 文書として保存していくことになるはずである。

一方追加課題とした二点についての成果はそれぞれ以下の通り:

(3) Jahnke Notesheet の電子データ化  
ノートシート 56 葉すべての電子データ化を完了し、その分析を通して現在保管されている形でのページの順序が本来並べられていたはずの順序と異なって乱れていることを実証し、正しい順序をできる限り復元した。

この研究成果は *Finnegans Wake* のテキスト生成過程研究史上大変重要な意義を持つはずだが、残念ながら Jahnke Notesheet は現時点で公刊不可の資料なので成果の公表は控えている。

(4) *Finnegans Wake* のテキスト生成過程研究用電子アーカイブの試作

この試作ではまず手持ちのエクセルのデータを Perl のデータ操作によって TEI (Text Encoding Initiative) のタグづけガイドラインに沿って XML 文書化する方法を確立し、次に Perl をつかってこの XML 文書から必要なデータを随意に取り出して文書閲覧や文書分析のためのさまざまなビューを生成する方法を習得した。この研究はまだ緒についたばかりだが、今後は出典発見や文書間の関係の分析を突き止める方法そのものをシミュ

レートしたビューを作ることによって、*Finnegans Wake*のテキスト生成過程研究を共同研究の形で一層効率化し加速化する方法を探っていきたいと考えている。さしあたり研究年度内にこの点について達成した成果としては、Google Books に出現する単語についての膨大なデータセットとして Google が公開している Google n-Gram Dataset を利用して、ノートブック VI. C. 2 に誤ったつづりで転記されている項目からもとの散逸したノートブック VI. D. 1 の項目を推測するための補助ツールを考案した。

#### 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 8 件)

- ① Robbert-Jan Henkes and Mikio Fuse, “Inside D1,” *Genetic Joyce Studies*, 査読有, 12, 2012, online.
- ② Mikio Fuse and Robbert-Jan Henkes, “Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 32,” *Genetic Joyce Studies*, 査読有, 12, 2012, online.
- ③ 扶瀬幹生、北本朝展, 「*Finnegans Wake* のテキスト生成過程研究のための TEI 準拠電子アーカイブの試作」、人文科学とコンピュータシンポジウム論文集: つながるデジタルアーカイブ—分野・組織・地域を超えて—, 7, 2012, 143-150.
- ④ Mikio Fuse and Robbert-Jan Henkes, “Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 10,” *Genetic Joyce Studies*, 査読有, 11, 2011, online.
- ⑤ Mikio Fuse and Robbert-Jan Henkes, “Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 14,” *Genetic Joyce Studies*, 査読有, 10, 2010, online.
- ⑥ Mikio Fuse, Robbert-Jan Henkes and Geert Lernout, “Further Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 3,” *Genetic Joyce Studies*, 査読有, 9, 2009, online.
- ⑦ Mikio Fuse and Robbert-Jan Henkes, “Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 16,” *Genetic Joyce Studies*, 査読有, 9, 2009, online.
- ⑧ Mikio Fuse and Robbert-Jan Henkes, “Emendations to the Transcription of *Finnegans Wake* Notebook VI. B. 5,” *Genetic Joyce Studies*, 査読有, 8, 2008, online.

[学会発表] (計 1 件)

① 扶瀬幹生、北本朝展, 「*Finnegans Wake* のテキスト生成過程研究のための TEI 準拠電子アーカイブの試作」、情報処理学会 人文学とコンピュータ研究会、2012. 12. 18、北海道大学。

#### 6. 研究組織

##### (1) 研究代表者

扶瀬 幹生 (FUSE MIKIO)  
聖心女子大学・文学部・教授  
研究者番号: 60262094

##### (2) 研究分担者

なし

##### (3) 連携研究者

なし