

デジタルコーパスを用いた データ駆動型の間テキスト性研究： 古代末期エジプトの二人の修道院長のコプト語書簡における コプト語訳聖書からの引用の探知と分析

宮川 創^{1,a)}

概要：本研究は、デジタルコーパスを用いたデータ駆動型による間テキスト性研究の一モデルを提示する。筆者らは、4世紀から5世紀にかけて生きた修道院長シェヌーテ、および、5世紀に生きた修道院長ベーサのコプト語による書簡のデジタル学術編集版を TEI XML 形式で作成した。そして、TRACER というテキスト・リユース（引用や引喩など）を探知するプログラムを用いて、シェヌーテとベーサの書簡コーパスと、ゲッティンゲン学術アカデミーから提供されたコプト語訳聖書コーパスとを比較・分析した。結果、先行研究では見つけることのできなかった聖書引用を多数発見した。本発表では、このような間テキスト性のデータ駆動型研究の手法とそれによって新しく見つかった聖書引用の特徴を考察する。

Data-Driven Research on Intertextuality Using Digital Corpora: Detection and Analysis of Quotations from Coptic Translation of the Bible in Coptic Epistles by Two Abbots in Late Antique Egypt

1. はじめに

1.1 間テキスト性

近年、あらゆるテキストを扱う学問分野において、間テキスト性 (intertextualité) という用語を用いた研究が盛んになっている。間テキスト性は、ブルガリア出身でフランスで活躍した哲学者ジュリア・クリステヴァ (Julia Kristeva) が、文学批評の文脈で初めて用いた用語である [1]。この概念はロシアの文学研究者ミハイル・バフチン (Mikhail Bakhtin) のカーニバル論、オーストリアの精神分析学者ジークムント・フロイト (Sigmund Freud) の精神分析学、そして、スイスの言語学者フェルディナン・ド・ソシュール (Ferdinand de Saussure) のアナグラム論の影響を受けて誕生したとされている [2]。

間テキスト性とは、クリステヴァの元々の文脈では、あらゆるテキストは、完全にオリジナルなものではなく、他のテキストから取られた「引用のモザイク」(mosaïque de citations) である性質を表す。その後、ジェラルド・ジュネット (Gérard Genette) などの哲学者や文学批評家により、間テキスト性の概念はより精緻化され、類型化されていった [3]。しかし、聖書学などフランス哲学や文学批評ではない分野においては、間テキスト性は、より一般的な意味で捉えられている。これらの分野では、間テキスト性研究の名のもと、引用、引喩、パラフレーズ、盗用など、あるテキストの著者が行った、別のテキストからの意識的・無意識的な行為が分析され、ソース (引用元)・テキストによるターゲット (引用先)・テキストへの影響が議論されている。

聖書文学学会 (Society of Biblical Literature) では、聖書本文批評学以外にも、様々な学問的切り口からの聖書研究 — 近年はデジタルヒューマニティーズのセッションも盛況である — の成果が発表されている。この学会は、アメリカ合衆国内の年次大会 (Annual Meeting) とアメリカ

¹ 京都大学大学院文学研究科 附属文化遺産学・人文知連携センター 助教

Kyoto University, Graduate School of Letters, Center for Studies of Cultural Heritage and Inter Humanities (CESCHI), Assistant Professor

^{a)} miyagawa.so.36u@kyoto-u.jp

合衆国外で行われる国際大会 (International Meeting) を両方とも一年に一度開催している。そのホームページ上には、毎年2つの大会の発表要旨集がウェブページの形で公開されており、そこにおける間テキスト性 —それがクリステヴァらが意図したものかどうかはさておき— の出現頻度から、聖書学における間テキスト性研究の「盛り上がり」、あるいは、「人気」の高低を調べることができる*1。国際大会は、毎回別の国で開催されており、近年では2016年のソウル大会などアジアでも開催されている*2。そのため、開催地によって、その開催国や開催地域の学術的関心の高まりが反映されやすく、発表要旨集における「間テキスト性」(intertextuality)への言及の頻度は年ごとにばらばらであった。それに対し、アメリカ合衆国内で開催される年次大会の発表要旨集においては一定の傾向が見て取れた。図1のグラフは、21世紀に入ってからの聖書文学学会年次大会の発表要旨集で言及された「間テキスト性」(intertextuality)と「間テキスト」(intertext)の頻度である。2000年代と比べ、2010年代では、この2つの用語の出現頻度は格段に増えている。

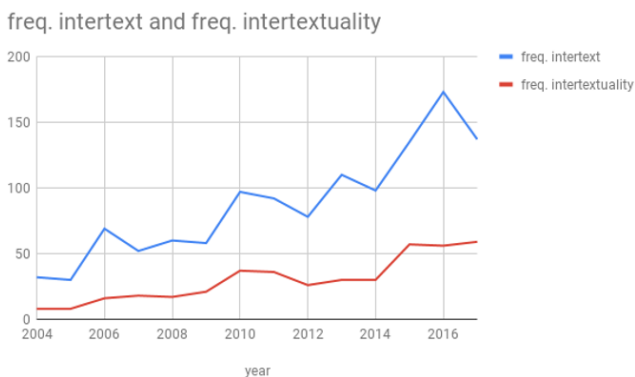


図1 聖書文学学会年次大会発表要旨集における intertextuality と intertext の出現頻度

Fig. 1 Frequency of intertextuality and intertext in abstract books of SBL Annual Meetings

図2は、Google Books N-Gram Viewer [5]によって生成されたものであり、Google Booksのテキスト全てを用いて、その語の出現頻度の全体語数に対する割合の変化を見ることができる。ここでも intertextuality と intertext への言及は年を経るたびに増えている。

1.2 テキスト・リユース研究と間テキスト性研究の邂逅

このように、2000年以降、間テキスト性研究が聖書学やその周辺分野で活発になってきた。一方、コンピュータ科学でもテキスト・リユース探知 (text reuse detection) [6]

*1 例えば、2018年の発表要旨集は [4]。

*2 2019年はインドのバンガロールで開催される予定であったが、開催地がローマの教皇庁立聖書研究所に変更された。

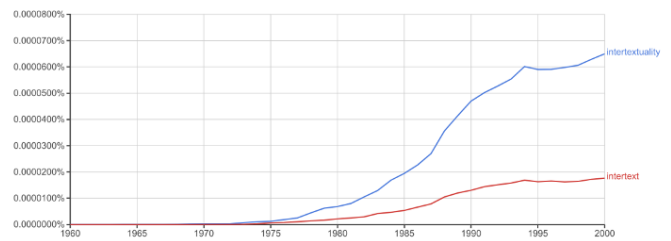


図2 Google Books N-Gram Viewerにおける intertextuality と intertext の分析

Fig. 2 Analysis on intertextuality and intertext by Google Books N-Gram Viewer

という分野があり、ここでは、引用、引喩、剽窃、言い回しなど、別のテキストで何らかの形で用いた既存のテキストを、コンピュータによって自動的に発見する装置の開発を目指している。この技術は、データマイニングからテキストマイニングの技術が発展し、その応用として、初期には実用的な事象、例えば、剽窃チェックなどに用いられた技術である。しかしながら、2000年代以降、隆盛している間テキスト性研究へ貢献できる技術として、聖書学の関連分野に応用されることが多くなっている。

ヨーロッパあるいは地中海世界の近世以前の文学では、引用する際に、出典を明記することが必ずしも行われていなかった。例えば、ラテン語で書かれた西洋古典文学では、出典が書かれず、さらには引用であることも示されていない引用が数多くある。そこで、Neil Coffeeが率いる Tesseract プロジェクト [7] では、ラテン語文学における明示されていない引用を、独自のテキスト・リユース探知技術を用いて探知した。その結果は、非常に注目すべきもので、古代ローマ白銀期の詩人ルカヌス (Lucanus) の『内乱』(De Bello Civili / Pharsalia) という作品の第1章において、作品が書かれてから約2,000年間、これまで先行研究が見つけられなかった、ウェルギリウス (Vergilius) の『アエネイス』(Aeneis) からの引用を46個発見した*3。

この Tesseract プロジェクトの成功とほぼ時を同じくして、ライプチヒ大学のコンピュータ科学者であった Marco Büchler は、自身の博士論文のプロジェクト [9] で、古典文学のためのテキスト・リユース探知プログラムである TRACER [10] を Java などを用いて開発した。このプログラムは、ライプチヒ大学の eAQUA [11] という古典ギリシア文学におけるテキスト・リユースを探知するプロジェクトで使用され、eTRACES [12] という後継のプロジェクトで用いられた。その後、Büchler は、ゲッティンゲン大学デジタルヒューマニティーズセンター (Göttingen Center for Digital Humanities; 略称 GCDH) へ異動すると同時に、ドイツ連邦研究教育省 (BMBF) の4年間160万ユーロの大型科学研究費を獲得して博士課程の学生やポスドク研究員

*3 この結果の詳細については [8] の p. 400 を見よ。

を雇い、2015年2月にeTRAPプロジェクトを始動させTRACERの更なる開発に努めた[13].

1.3 テキスト・リユース探知技術のコプト語文献における間テキスト性研究への応用

同時期の2015年6月にドイツ研究振興協会によってゲッティンゲン大学に設置された特別研究領域/共同研究センター1136「古代から中世及び古典イスラム期にかけての地中海圏とその周辺文化における教育と宗教」*4は、アトリペのシェヌーテという紀元後4~5世紀の修道院長のコプト語書簡集の1つである『第6カノン』とその弟子ベーサの『書簡と説教集』のTEI XML [15]によるデジタル学術編集版を作成し、聖書との間テキスト性を分析するというサブプロジェクトを始めた。コプト語は、古代エジプト語の直接の末裔で、ギリシア文字24字にエジプト民衆文字由来の6~7文字を補足したコプト文字で書かれたエジプト語である。古代末期キリスト教やキリスト教初期修道制の文献、新約聖書の最古層の翻訳、マニ教文献、グノーシス主義文献など宗教学的に貴重な資料がコプト語で多く残っている。コプト語には様々な方言があるが、本稿の研究対象である書簡およびコプト語訳聖書は、サヒド方言(サイド方言)で書かれている。2015年10月に筆者はこのプロジェクトの学術研究員(Wissenschaftlicher Mitarbeiter)に着任し、当該文献のデジタル学術編集版の作成に携わったほか、聖書からの間テキスト性を分析するため、eTRAPチームと共同でテキスト・リユース探知を行った。その後2020年6月30日まで本プロジェクトは続いた。

2. 方法論

2.1 コーパス

本プロジェクトの分析対象となる2つのコーパスのうちの1つのシェヌーテ『第6カノン』には6つの写本の諸ページが、もう1つのベーサの『書簡と説教集』には2つの写本の諸ページが残っている。これらの写本はページ毎に分割されて大部分はフランス国立図書館、ナポリ国立図書館、大英図書館など欧米の図書館に、一部はエジプト・カイロのコプト博物館に所蔵されている。シェヌーテ『第6カノン』の諸写本の多くのページは、100年以上前にÉmile Amélineau [16][17]とJohannes Leipoldt [18][19]による翻刻版が出版されており、ベーサ『書簡と説教集』は約65年前にKarl-Heinz Kuhn [20]による翻刻版が出版されている。しかし、殊にAmélineauの翻刻は非常に誤りが

*4 ドイツ語: Sonderforschungsbereich 1136 "Bildung und Religion in Kulturen des Mittelmeerraums und seiner Umwelt von der Antike bis zum Mittelalter und zum Klassischen Islam", 英語: Collaborative Research Centre 1136 "Education and Religion in Cultures of the Mediterranean and Its Environment from Ancient to Medieval Times and to the Classical Islam"[14].

多く、写本と付き合わせて修正する必要があった。

本研究で用いられた工程を以下の図3で示す。以下、はじめのOCRの工程から述べていく。

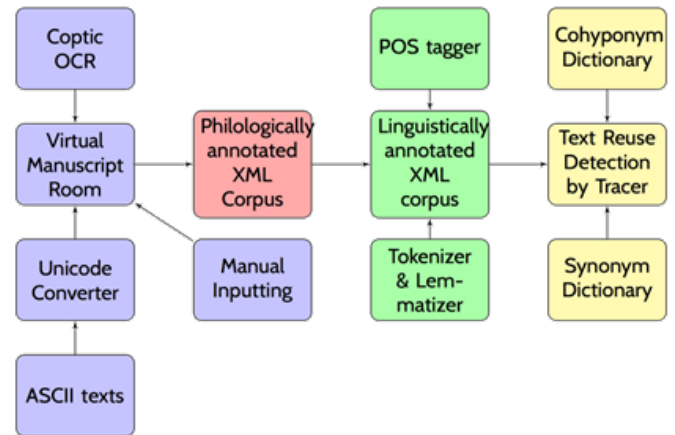


図3 本研究のワークフロー

Fig. 3 Workflow of this study

2.2 コプト語 OCR と Unicode コンバータ

これらの、活版印刷された古い翻刻版を自動でデジタルテキストにするために、深層学習を用いたOCR(光学文字認識)プログラムであるOCRopy*5をこれらの翻刻版の一部でトレーニングさせ、最終的にはこれらの翻刻版の全体をデジタルテキスト(Unicode)化した。図4は、初期のトレーニング時のトレーニング回数と、トレーニングデータとは全く異なる別の翻刻版における精度のグラフ[23]であり、ほぼ90%以上の精度を保っている*6。

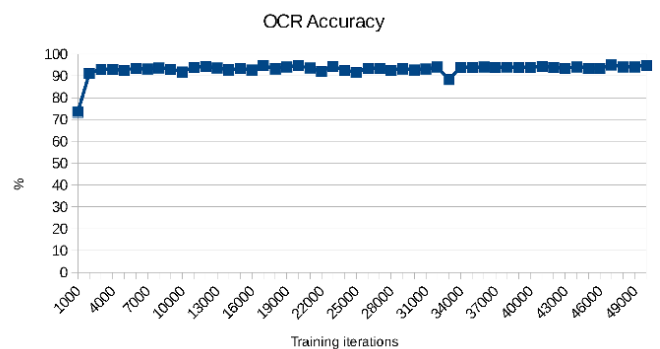


図4 OCRopyをコプト語テキストに用いた際のトレーニング回数とトレーニングデータとは異なるテキストを読み取った際の精度

Fig. 4 Training iteration and accuracy of OCRopy

ほかに、過去、コプト語文献学者たちがASCIIフォントで

*5 C++で書かれたOCRopusの、CLSTMを用いたPython版である[21].

*6 コプト語テキストにおける人工ニューラルネットワークを用いたOCRに関しては[24]を見よ.

『第6カノン』の一部の翻刻を行ったものがある。これらをUnicodeに直すために、Caroline Schroeder, Amir Zeldesと筆者が開発したPerlによるUnicodeコンバータ [22] を用いてこれらのUnicode化を行った。

OCRもしくはUnicodeコンバータを通して得られたUnicodeによる翻刻は、後に述べるVirtual Manuscript Room (VMR) というウェブアプリケーションにアップロードし、写本の写真と突き合わせて、誤りの修正を行った。また、既存の翻刻がないページは、VMR上で写本の写真を元に一から手作業でデジタル翻刻した。

2.3 Virtual Manuscript Room を用いたデジタル翻刻とEpidocによる出力

Virtual Manuscript Room (VMR) [25] は、CrossWire Bible Society [26] という無料の聖書分析ソフトを多数作っていた協会の主要プログラマーだったTroy Griffithsが、ミュンスター大学新約聖書本文研究所 (Instituts für Neutestamentliche Textforschung) とバーミンガム大学の本文研究・電子編集研究所 (Institute for Textual Scholarship and Electronic Editing) において、新約聖書諸写本のデジタル翻刻と校訂のために、マルキオン研究者でDHにも通じていたUlrich Schmidとともに開発したウェブアプリケーションである。はじめは、様々な聖書翻訳の底本になっているNestle-Aland版ギリシア語新約聖書の拡大版であるEditio Critica Maior (ECM) を編纂するためにこのVMRが用いられた。

VMRでは、簡潔なGUIエディタ (図5) によって、アップロードした写本の写真と見比べながら写本の翻刻を行い、様々なマークアップを施すことができる。作成された翻刻データはTEI XMLのサブセットであるEpidoc [27] 形式やHTMLで出力することができる。画像は直接アップロードすることもできるが、他の図書館や研究機関がIIIF Manifestを公開している場合、それを読み込んで表示することもできる。プラットフォームとしてLiferay [28], エディタにはTiny MCE [29] が用いられ、共同作業のバージョン管理のためにGitが、データの格納にMySQLが用いられている。

Griffithsは、ゲッティンゲン学術アカデミーのコプト語旧約聖書デジタル・エディションのプロジェクトのために、コプト語文献の翻刻に特化したVMRを構築した [30]。このVMRを用いて、シェヌーテ『第6カノン』とペーサ『書簡と説教集』の諸写本の翻刻を写真と付き合わせて修正し、Epidoc形式のTEI XMLファイルとして出力した。

2.4 Coptic NLP Service を用いたコプト語形態素・統語解析

Coptic NLP Service [31] は、Coptic SCRIPTORIUM [32][33] という全米人文科学基金 (NEH) のコプト語タグ付

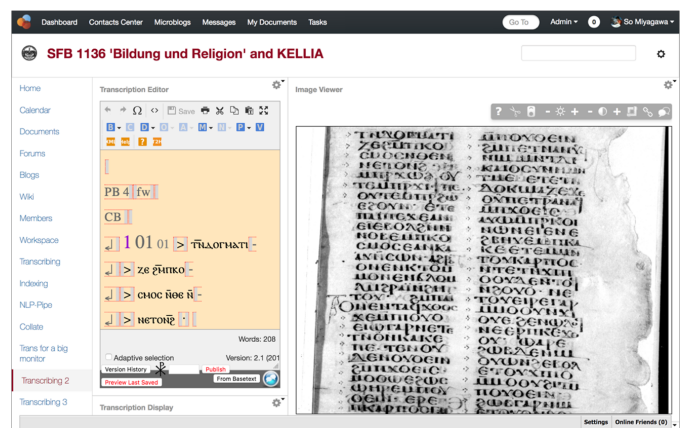


図5 VMRのエディタ画面

Fig. 5 Editor on VMR

き多層コーパスプロジェクトであり、2013年に始まり現在も続いている。コプト語のコーパスを、クエリ言語による単語・構文検索ができるANNIS (図6下; [37]) というプラットフォームで公開しているほか、Universal Dependencies [38] による統語情報をタグ付けしたツリーバンクもArborator [34] で表示させて提供している (図6上)。Coptic SCRIPTORIUM プロジェクトは、全米人文科学基金とドイツ研究振興協会のジョイント・プロジェクトKELLIAの一員であり、このKELLIAプロジェクト [35] のもと、入力したコプト語の形態素および統語情報をGUI上で解析できるCoptic NLP Serviceが作成された。

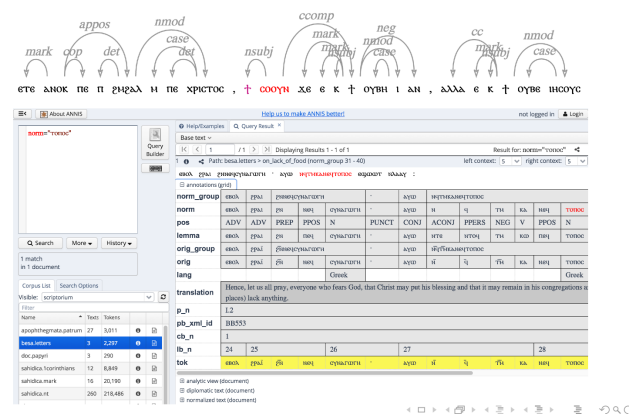


図6 Coptic SCRIPTORIUM のコーパスの、ANNISによる表示

Fig. 6 Coptic SCRIPTORIUM corpus displayed on ANNIS

このCoptic NLP Serviceは、コプト語の形態素解析をするようトレーニングされたTreeTagger [36] とUniversal Dependenciesによるコプト語の統語解析をするようにトレーニングされたMaltParser [39] が組み合わせられており、レンマ・品詞・統語情報が含まれたSGMLファイルを作成する*7。このサービスではTEI XMLファイルのタグが入っていても問題なく解析される。

*7 この作業の詳細や引用情報をも含めたSGMLの拡張に関しては、[40]を見よ。

筆者らは、VMRで作成したシェヌーテとベーサのデジタル学術編集版のTEI XMLファイルをCoptic NLP Serviceにかけ、形態素・統語解析されたSGMLファイルを作成し、それをXMLファイルに再変換して、TRACERによるテキスト・リユース探知に備えた。

なお、本研究は、シェヌーテとベーサの聖書引用の探知が目的であるため、コプト語訳聖書の同様のXMLファイルが必要であった。コプト語訳聖書のデジタルデータはゲッティンゲン学術アカデミーから提供され、同様の手段で形態素・統語解析した。これらのTEI XMLファイルを用いて、2人の修道院長の著作と聖書の間のテキスト・リユースをTRACERで探知した。

2.5 テキスト・リユース探知プログラム TRACER

TRACERはJavaで書かれたプログラムであり、700以上のアルゴリズムを持つ。それぞれのアルゴリズムのパラメータは、XMLで書かれたコンフィギュレーションファイル(config.xml)を修正することで調節できる。現在までに、ラテン語、古代ギリシア語、チベット語、ヘブライ語、アラビア語、ドイツ語、英語においてTRACERは使用された。本研究で、コプト語がここに追加されたことになる。

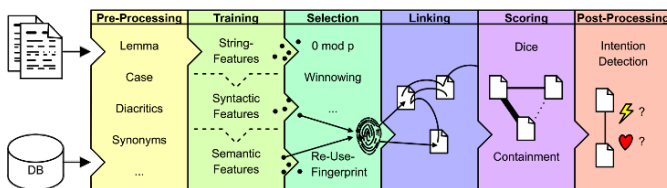


図7 TRACERの工程
 Fig. 7 Process of TRACER

TRACERには6つのレベルの工程があり、図7 [41]で示されている。レベル1は、Pre-processing (前処理)であり、日本語や中国語などの分ち書きをしない文字 (scriptura continua) では、ここで単語間にスペースを入れる (word segmentation)。また、ダイアクリティカルマークや補助記号などが可変的な言語では、これらの記号を取る (normalization)。このようにして後の工程の精度を高める下準備をする。

この時点で、WordNetやBabelNetなど、同義語、下位語、上位語など単語の意味ネットワークが記述されている概念辞書のファイルとリンクさせ、西洋の古代・中世の引用で生じやすい同義語や共-下位語*8による語彙の置き換えにも対応するようにする。本研究では、筆者がLaura Slaughter (オスロ大学) およびLuís Morgado da Costa (南

*8 同じ上位語 (hypernym) を共有する下位語 (hyponym) 同士の関係を共-下位語 (co-hyponym) と呼ぶ。シェヌーテとベーサの著作における引用は多くは、引用であることが明示されておらず、かつ、引用の一部の語彙が同義語や共-下位語に変更されていることが多い。

洋理工大学) らと共同開発したコプト語 WordNet[42]を用いた。

そのほか、このレベル1で、Coptic NLP Serviceでタグ付けしたレンマや品詞の情報を読み込む。この情報は、後のレベル3: Selectionで発揮される。

レベル2は、Featuring/Training (素性化/トレーニング)であり、ここでは、2つのテキストをTRACERが比較する際の単位を設定し、その単位ごとに文字列をテキストから抽出する。パラメータの設定で、単語を単位とする方法、バイグラムを単位とする方法、トリグラムを単位とする方法、それ以上のグラムを単位とする方法がある。

レベル3は、Selection (選択)であり、冠詞や代名詞、前置詞など、頻度が多いがテキスト・リユースの探知に直接影響を及ぼさないストップ・ワードなどを分析の対象から除外する。

レベル4は、Linking (リンク付け)であり、これまでに単位毎に文字列のユニットが抽出されてきた2つのテキストを比較し、類似している、つまり、多くの文字列のユニットがオーバーラップしている部分を導き出す。結果が少ない場合はMoving Window法という措置をとる。これは、テキスト上にさらに10から15グラムの単位のウィンドウをかぶせ、それを1ユニットずつずらしていき、テキストから複数のウィンドウを生成して、2つのテキストをウィンドウを基準に比較し、類似度を数値化していく手法である。これにより、句レベルなどより小さなテキスト・リユースを見つけることができる。本研究では、このMoving Window法を用いた。

レベル5のScoring (スコア付け)では、これまでにリンク付けされたテキストのそれぞれの箇所の類似度を数値化する。数値化の際にabsolute overlapとweighted overlapという2つの方法を用いる。absolute overlapは、2つのテキストの比較で13以上のユニットが一致したテキスト・リユース候補で用いられ、一致した要素数を値とする。weighted overlapは、テキスト・リユース候補におけるユニットの共有の比率をパーセンテージで数値化する。

レベル6は、Visualization (視覚化)であり、これは、Stefan Jänickeが開発した、TRACER用の視覚化プログラムTRAViz[43]を用いたものである。視覚化によって生成されるウェブページにはJavaScriptによる動的な3つのビューがある。すなわち、ドット・プロット・ビュー (dot-plot view; 図8)、テキスト・リユース・アライメント・ビュー (text re-use alignment view; 図9)、および、パラレル・テキスト・ビュー (parallel text view; 図10)である。

ドット・プロット・ビュー (図8)では、2つのテキストをそれぞれx軸とy軸に配置し、2つのテキストの中でテキスト・リユースの可能性のある箇所を小円で表す。小円は、黄色いほど2つのテキストのそれぞれの箇所の類似性がより低く、緑になるほど高くなる。

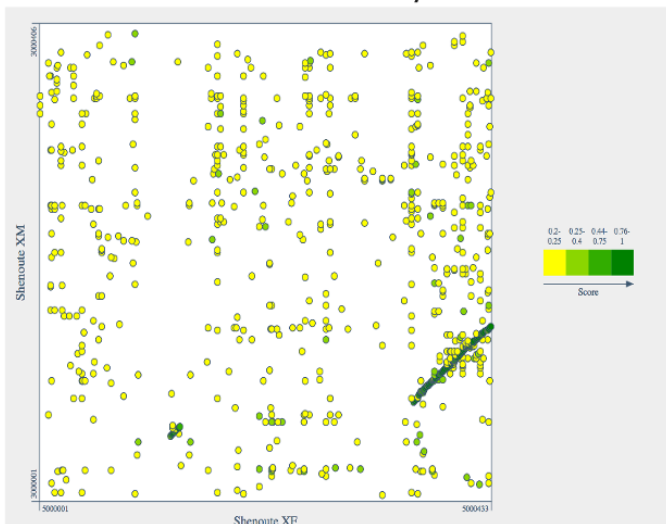


図 8 TRAViz によるドット・プロット・ビュー
Fig. 8 Dot-plot view by TRAViz

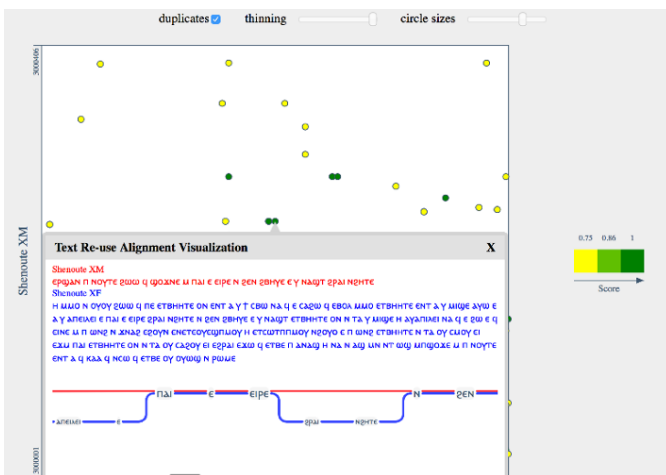


図 9 TRAViz によるテキスト・リユース・アライメント・ビュー
Fig. 9 Text re-use alignment view by TRAViz

ドット・プロット・ビューの小円をクリックすると、テキスト・リユース・アライメント・ビュー (図9) がポップアップする。ここでは、その箇所の2つのテキストを平行線で示すが、語が一致している箇所では、それらの平行線が曲がり、1つに重なる。

最後に、パラレル・テキスト・ビューであるが、ここでは、2つのテキストの全体がそれぞれミニ・ウィンドウ内で表示され、類似性が高く、テキスト・リユースである可能性がある箇所では、それら2つの地点が黄色や緑の線で結ばれる。色はドット・プロット・ビューと同じく、黄色であるほど類似性が低く、緑であるほど類似性が高い。黄色や緑の線をクリックすれば、それが結んでいる類似度が高い2つのテキストの地点が平行で表示され、類似部分は青くハイライトされる。

これら3つのビューを用いて、TRACERが導き出したテキスト・リユースの候補が本当に引用や引喩であるのか、



図 10 TRAViz によるパラレル・テキスト・ビュー
Fig. 10 Parallel text view by TRAViz

精査した。以下がその結果である。

3. 結果

上述の工程を経て、TRACERが探知したテキスト・リユース候補の数は以下の表1の通りである。

	シェヌーテ『第6カノン』	ペーサ『書簡と説教集』
聖書	8,414	13,835

表 1 コプト語訳聖書からのテキスト・リユース候補数
Table 1 Number of text re-use candidates from the Coptic Bible

TRACERによって見つかったテキスト・リユースの候補は合計で22,249であり、これを全て前節の3つのビューで調べていくには、膨大な時間と労力がかかることが予想された。そのため、同時代のコプト語著述家によって最もよく引用されたことが知られている、聖書中の書物の1つである『詩編』からのテキスト・リユース候補のみを対象として分析することにした。以下の表2は、シェヌーテ『第6カノン』とペーサ『書簡と説教集』における『詩編』からのテキスト・リユースの候補数である。

	シェヌーテ『第6カノン』	ペーサ『書簡と説教集』
『詩編』	423	629

表 2 『詩編』からのテキスト・リユース候補数
Table 2 Number of text re-use candidates from the Psalms

これらのテキスト・リユースの候補数は合計で1,052であり、前節で示したTRAVizによる3つのビューを用いて、それらがこういった種類のテキスト・リユースであるかを分析した。そして、その後それらのテキスト・リユースが先行研究において言及されているかを調べた。その結果、以下の表3が得られた。

表3はKarl-Heinz Kuhn [20], Émile Amélineau [16][17],

	シェヌーテ『第6カノン』	ペーサ『書簡と説教集』
『詩編』	14	6

表 3 先行研究には見られなかった引用

Table 3 Quotations which were not attested in previous literature

Hermann Wiesmann [44][45] などの先行研究が見つかることができなかった引用である。とくに Kuhn [20] は、厳格な文献学者として知られ、非常に多くの『詩編』からの引用を、ペーサの『書簡と説教集』の中から見出したが、今回 TRACER が探知した6つの引用を、Kuhn は見出せなかった。このように、TRACER の有用性が、今回、コプト語においても示された。これら TRACER によって新しく見つかった引用は、シェヌーテやペーサが聖書を引用する際によく用いる *nthe etsêh* 「書かれてあるように」*9 などの引用を示唆する導入句が用いられておらず、また、語彙や語順の変更が行われているものが多かった。これらの要因から先行研究は、これらの TRACER が新しく見つけた引用を発見することができなかったものだと思う。

4. 今後の課題

以上、データ駆動型の間テキスト性研究の一モデルを提示し、その結果について分析した。

この分析の最後に、先行研究が見つけれず、TRACER にもみつけられた引用を示し、本研究で用いた方法の有用性を示した。しかし、TRACER は、引用以外にも、聖書で何度も用いられている言い回しや、聖書からの引喩を、シェヌーテ『第6カノン』とペーサ『書簡と説教集』の中で大量に探知している。今後は、引用以外にこれらの分析も行う必要がある。

そして、先行研究が既に発見しているテキスト・リユースと今回 TRACER が新しく発見したテキスト・リユースをもとに、シェヌーテとペーサの、聖書からのテキスト・リユースにおける文体差について分析することも課題である。特に、引用における語句や語順の変異の度合いをシェヌーテとペーサで比べることで、どちらがより原典に忠実か、またはどちらが、原典を文脈に合わせて自由に改変しているかを知ることができる。このような分析は、著者不明のコプト語修道文献の著者を同定する際の書体上のヒントとなることが予想される。

そのほか、先行研究が発見できず、TRACER によってのみ発見された引用は、引用の導入句がなく、語彙や語順

*9 「詩篇に書かれてあるように」など聖書の書名を明示したり、「預言者が言っているように」など聖書の預言書や詩篇に書かれていることを示唆することも時々あるが、多くの場合、このような導入句においてその出典は言及されない。そもそも、こうした導入句を用いず、引用であることを明示しない引用はシェヌーテおよびペーサにおいて非常に多い。

の変更がなされたものが多かったが、逆に先行研究が発見できて、TRACER が発見できなかった引用も数多くあった。このような先行研究、すなわち、人間のみが発見できたテキスト・リユースの特徴を分析することも TRACER の機能の向上に繋がる、今後の大きな課題の1つである。

謝辞

本研究は、ドイツ研究振興協会 (Deutsche Forschungsgemeinschaft; 略称 DFG) の特別研究領域 (Sonderforschungsbereich; 略称 SFB; 英語名称 Collaborative Research Center「共同研究センター」) 1136「古代から中世及び古典イスラム期にかけての地中海圏とその周辺の文化における教育と宗教」(Sonderforschungsbereich 1136 “Bildung und Religion in Kulturen des Mittelmeerraums und seiner Umwelt von der Antike bis zum Mittelalter und zum Klassischen Islam”) のサブプロジェクト B05「コプト語を用いるエジプトの古代末期キリスト教における聖書解釈と教育伝統: シェヌーテ『第6カノン』」(Schriftauslegung und Bildungstraditionen im koptischsprachigen ägyptischen Christentum der Spätantike: Schenute, *Kanon* 6) において遂行された。本研究の一部は科研費 JP21K00537, JP20K21975, JP15J05370 を受けて遂行されたものを含んでいる。最後に、本稿に有益なコメントを下された永崎研宣先生 (人文情報学研究所) と大向一輝先生 (東京大学) に感謝を表したい。

参考文献

- [1] Kristeva, J.: *Semeiotikê : recherches pour une sémanalyse*, Éditions du Seuil, Paris (1969).
- [2] 岩本一: 間テキスト性 —その展開と関連性について—, *dialogos*, No. 1, pp. 39–57, (2001).
- [3] Genette, G.: *Palimpsestes : la littérature au second degré*, Éditions du Seuil, Paris (1992).
- [4] Society of Biblical Literature: Meeting Abstracts, 2018 Annual Meeting, Denver, CO, https://www.sbl-site.org/meetings/Congresses_Abstracts.aspx?MeetingId=33 (accessed 2021-04-20).
- [5] Google: Google Books Ngram Viewer, <https://books.google.com/ngrams> (accessed 2021-04-21).
- [6] Greta, F, Franzini, E., Büchler, M.: Historical Text Re-use: What Is It? (2016), <http://www.etrapp.eu/historical-text-re-use/> (accessed 2019-04-11).
- [7] Coffee, N., Koenig, J.-P., Poornima, S., Forstall, C. W., Ossewaarde, R., Jacobson, S. L.: The Tesseræ Project: Intertextual Analysis of Latin Poetry, *Literary and Linguistic Computing*, Vol. 28, No. 2, pp. 221–228 (2013).
- [8] Coffee, N., Koenig, J.-P., Poornima, S., Ossewaarde, R., Forstall, C. W., Jacobson, S. L.: Intertextuality in the Digital Age, *Transactions and Proceedings of the American Philological Association*, Vol. 142, No. 2, pp. 383–422 (2012), available at https://www.researchgate.net/publication/259574933_Intertextuality_in_the_Digital_Age (accessed 2021-04-22).
- [9] Büchler, M.: Informationstechnische Aspekte des Historical Text Re-use, Dissertation, Universität

- Leipzig (2013), available at <https://ul.qucosa.de/api/qucosa%3A11877/attachment/ATT-0/> (accessed 2021-04-22).
- [10] eTRAP: TRACER, <https://www.etrapp.eu/research/tracer/> (accessed 2021-04-22).
- [11] Buechler, M., Heyer G., Gründer, S.: eAQUA—Bringing modern Text Mining approaches to two thousand years old ancient texts, *Proceedings of e-Humanities—An Emerging Discipline, workshop at the 4th IEEE International Conference on e-Science* (2008), available at <http://asv.informatik.uni-leipzig.de/publication/file/123/IEEE09.pdf> (accessed 2021-04-20).
- [12] Göttingen Center for Digital Humanities, eTRACES, <https://www.gcdh.de/en/research/projects/etraces/> (accessed 2021-04-21).
- [13] eTRAP, <https://www.etrapp.eu/> (accessed 2021-04-22).
- [14] Sonderforschungsbereich 1136 “Bildung und Religion in Kulturen des Mittelmeerraums und seiner Umwelt von der Antike bis zum Mittelalter und zum Klassischen Islam,” <http://www.sfb1136.uni-goettingen.de/> (accessed 2021-04-21).
- [15] Text Encoding Initiative, P5: Guidelines for Electronic Text Encoding and Interchange, <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (accessed 2021-04-20).
- [16] Amélineau, É.: *Œuvres de Schenoudi, Texte Copte et Traduction Française*, Vol. 1, Ernest Leroux, Paris (1907).
- [17] Amélineau, É.: *Œuvres de Schenoudi, Texte Copte et Traduction Française*, Vol. 2, Ernest Leroux, Paris (1914).
- [18] Leipoldt, J.: *Sinuthii Archimandritae Vita et Opera Omnia III*, Corpus Scriptorum Christianorum Orientalium 42, Scriptorum Coptici 4, Imprimerie Nationale, Paris (1908).
- [19] Leipoldt, J.: *Sinuthii Archimandritae Vita et Opera Omnia IV*, Corpus Scriptorum Christianorum Orientalium 73, Scriptorum Coptici 5, Imprimerie Nationale, Paris (1913).
- [20] Kuhn, K. H.: *Letters and Sermons of Besa*, Corpus Scriptorum Christianorum Orientalium 157, Scriptorum Coptici 21, Imprimerie Orientaliste L.Durbeq, Louvain (1956).
- [21] Breuel, T., et al.: OCRopy, <https://github.com/ocropus/ocropy> (accessed 2021-04-21).
- [22] Zeldes, A., Schroeder, C. T., Miyagawa, S.: `recode_coptic.pl`, https://github.com/CopticScriptorium/converters/blob/master/recode_coptic.pl (accessed 2021-04-21).
- [23] Miyagawa, S., Bulert, K., Büchler, M.: Utilization of Common OCR Tools for Typeset Coptic Texts, Abstract, 2nd International Conference on Digital Access to Textual Cultural Heritage, DATECH 2017, Göttingen, Germany, June 1-2, 2017 (2017).
- [24] Miyagawa, S., Bulert, K., Büchler, M., Behlmer, H.: Optical character recognition of typeset Coptic text with neural networks, *Digital Scholarship in the Humanities*, Vol. 34, No. suppl.1, pp. i135–i141, DOI:10.1093/llc/fqz023 (2019).
- [25] Griffiths, T., Virtual Manuscript Room Collaborative Research Environment (VMR CRE), <https://vmrcrcre.org/> (accessed 2021-04-20).
- [26] CrossWire Bible Society, <https://crosswire.org/> (accessed 2021-04-22).
- [27] Elliott T., Bodard, G., Cayless, H., et al.: EpiDoc: Epigraphic Documents in TEI XML, <https://epidoc.stoa.org/gl/latest/> (accessed 2021-04-22).
- [28] Liferay, <https://www.liferay.com/> (accessed 2021-04-21).
- [29] TinyMCE, <https://www.tiny.cloud/get-tiny/> (accessed 2021-04-21).
- [30] Akademie der Wissenschaften zu Göttingen: Digital Edition of the Coptic Old Testament, <http://coptot.manuscriptroom.com/> (accessed 2021-04-22).
- [31] Coptic NLP Service, <https://corpling.uis.georgetown.edu/coptic-nlp/> (accessed 2021-04-21).
- [32] Coptic SCRIPTORIUM, <https://copticcriptorium.org/> (accessed 2021-04-21).
- [33] Schroeder, C. T., Zeldes, A.: Raiders of the Lost Corpus, *Digital Humanities Quarterly*, Vol. 10, No. 2 (2016), available at <http://www.digitalhumanities.org/dhq/vol/10/2/000247/000247.html> (accessed 2021-04-21).
- [34] Gerdes, K.: Arborator, <https://arborator.ilpga.fr/> (accessed 2021-04-22).
- [35] KELLIA (the Koptische/Coptic Electronic Language and Literature International Alliance), <https://kellia.uni-goettingen.de/> (accessed 2021-04-22).
- [36] Schmid, H.: TreeTagger — a part-of-speech tagger for many languages, <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (accessed 2021-04-21).
- [37] corpus-tools.org: ANNIS, <https://corpus-tools.org/annis/> (accessed 2021-04-21).
- [38] Universal Dependencies, <https://universaldependencies.org/> (accessed 2021-04-21).
- [39] Hall, J., Nilsson J., Nivre, J.: MaltParser, <http://www.maltparser.org/> (accessed 2021-04-21).
- [40] Miyagawa, S., Zeldes, A., Büchler, M., Behlmer, H., Griffiths, T.: Building Linguistically and Intertextually Tagged Coptic Corpora with Open Source Tools, in Suzuki, C. (ed.): *Proceedings of the 8th Conference of Japanese Association for Digital Humanities*, pp. 139–41, Center for Open Data in the Humanities, Tokyo (2018).
- [41] eTRAP: Configuration, TRACER, <https://gfranzini.gitbooks.io/tracer/content/manual/configuration/> (accessed 2021-04-21).
- [42] Slaughter, L., Morgado da Costa, L., Miyagawa, S., Buchler, M., Zeldes, A., Lundhaug, H., Behlmer, H.: The Making of Coptic Wordnet, in Fellbaum, C., Vossen, P., Rudnicka, E., Maziarz, M., Piasecki, M. (eds.): *Proceedings of the Tenth Global Wordnet Conference*, pp. 166–175, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2019), available at <https://www.aclweb.org/anthology/2019.gwc-1.21.pdf> (accessed 2021-04-21).
- [43] Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S., Scheuermann, G.: TRAViz: A Visualization for Variant Graphs, *Digital Scholarship in the Humanities*, Vol. 30, No. suppl.1, pp. i83–i99, DOI: 10.1093/llc/fqv049 (2015).
- [44] Wiesmann, H. (interp.): *Sinuthii Archimandritae Vita et Opera Omnia III*, E Typographeo Reipublicae, Paris (1931).
- [45] Wiesmann, H. (interp.): *Sinuthii Archimandritae Vita et Opera Omnia IV*, E Typographeo Reipublicae, Paris (1936).