様　式　C－19、F－19、Z－19　(共通)

科学研究費助成事業　　研究成果報告書

平成 27 年　 6 月 10 日現在

機関番号：12102
研究種目：基盤研究(C)
研究期間：2012～2014
課題番号：24520452

研究課題名（和文）高精細画像とXMLデータを用いた古代エジプト語文書の言語記述アーカイブズの構築

研究課題名（英文）Construction of a Database Archive for Egyptian Hieratic Using High-Resolution Digital Images and XML Format Data

研究代表者

永井　正勝（NAGAI, Masakatsu）

筑波大学・人文社会系・助教

研究者番号：70578369

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本研究では、古代エジプト語の神官文字パピルスBM10221（大英博物館所蔵）を対象に、神官文字の高精細画像と言語情報（文字や語の情報）とをリンクさせたWEB型データベースを英文で作成した。本データベースのメイン画面では、神官文字画像の任意の箇所にカーソルを合わせると、文字情報と単語情報がポップアップ画面に示されるように実装した。これにより、原資料の写真を参照しながら、テキストを読み解くことができる。また、検索画面では、言語情報の各要素から検索を行うと、検索結果が原資料の写真と共に示される。出力画面からは、TEIならびにLODに準拠したタグ付きテキストをXML形式で出力させることができる。

研究成果の概要（英文）：We constructed a Web-based database for Egyptian Hieratic Papyrus EA10221, owned by the British Museum, which contains high-resolution digital images and linguistic information. In the gallery window of this database, if the user places the cursor on a place that he/she needs to know about, then linguistic information such as the code, phoneme, and function of the sign, and also such things as the actual written form, the dictionary form,the grammatical category, and meaning of the word will be shown in the pop-up windows with pictures of the target signs. In the search window, users could search for signs or words from linguistic information. The results will be shown with pictures of the signs from an original text. In the output window, users could download data with annotations in XML format based on TEI or LOD.

研究分野：言語学

キーワード：古代エジプト語　アフロ=アジア諸語　写本　表記法　字形　画像データベース　Text Encoding Initiative　Linked Open Data (LOD)

１．研究開始当初の背景

本研究の背景となるのは、(1)言語学における文献資料の扱いならびに(2)エジプト学における神官文字資料の扱いに関する疑問である。以下に、それぞれの分野における問題点を指摘する。

(1)言語学における文献資料の扱いに関する問題点

文献資料とりわけ写本には、書字材料の種類と制作方法、保存状態、インクの種類や色、捺印、挿絵、文書の書式、筆記者、宛名、受取人、墨継ぎの箇所、書字方向、字形など、いわば言外の情報が豊富に含まれている。言語学における研究では、上記のような言外の情報を捨象したうえで、言語テキストを抽出し、言語の体系や構造の記述が目指される。しかしながら、通常の言語研究で捨象される言外の情報が、言語テキストを理解するための鍵となることもある。また、字形や書字方向に関する情報は、表記法の分析とあわせ、文字学・文字論の領域として言語学の内部に立ち位置を設ける必要があるが、現在の言語学研究では、文字そのものに関する関心が低い。このような問題意識のもと、本研究では古代エジプト語の写本の字形情報に着目することとした。

言語テキストの作成について、先に言外の情報を切り捨てると述べたが、しかしながら、言語テキストの作成は、情報を切り捨てるという単純な作業のみで成り立っているわけではない。むしろ、言語テキストの抽出とは、原資料の文字と対峙し、その文字がいったいどのような文字であるのか、あるいはそこにどのような形態素や言語形式が隠れているのかということを、複数の選択肢の中から逐一判断しつつ、現代の学者が理解可能な状態に資料を変換させるという行為であると言える。つまり、原資料に対する解釈の結果が、言語テキストとして存在しているのである。言語テキストは、ある種の転写で表示されることになるが、その結果として、原資料と言語テキストとの間に乖離が生じることになる。この乖離を補うべく、従来の研究では、写真図版がどちらかと言えば補助的に用意されていた。しかしながら、本来参照すべき資料は、原資料そのものであり、学者が作成した言語テキストではない。このような言語テキストと原資料とのねじれた関係を是正させるべく、本研究では原資料の写真を用いたデータベースを構想した。ただしこれは、言語テキストを用いないということではなく、言語テキストを原資料の写真と並置させることにより、言語テキストの価値を相対化させることを意図している。

(2)エジプト学における神官文字資料の扱いに関する問題点

古代エジプト語は紀元前 3100 年頃に文字化され、紀元後 18 世紀頃まで使用された言語であり、アフロ=アジア諸語に属する。古代エジプト語は、原エジプト文字、聖刻文字、神官文字、民衆文字、コプト文字 (ギリシャ文字) によって表記されたが、それぞれの文字の使用は時代とテキストのジャンルに基づいて異なっていた。これらの文字のなかで、現在の研究において、異色な位置づけにあるのが神官文字である。というのも、それぞれの文字によって記されたテキストはそれぞれの文字で示されるのが通常であるはずだが、エジプト学では神官文字のテキストのみが神官文字で示されず、聖刻文字に転写された「聖刻文字翻字」が「資料」として頻繁に使用されるからである。

エジプト学の世界では、聖刻文字を筆記体に崩したものが神官文字であるとの認識を持つ者が多いが、聖刻文字と神官文字はともに原エジプト文字から派生したものであって、神官文字は聖刻文字の筆記体ではない。実際、聖刻文字にしか認められない文字や神官文字にしか認められない文字があり、聖刻文字と神官文字は十全な意味で１対１に対応するわけではない。それにもかかわらず、エジプト学では聖刻文字翻字が「資料」として受け入れられており、神官文字を読むことのできない、あるいは神官文字を参照していない学者達が、聖刻文字翻字を利用して各種の分析を行っているのである 。いうまでもなく、聖刻文字翻字には、それを作成した人の判断が多分に入り込んでおり、文字判読の結果が読み手によって異なることが少なくない。つまり、聖刻文字翻字はある種の言語テキストであり、原資料そのものではない。

エジプト学におけるこのような状況、つまり、聖刻文字翻字のみを参照しているという状況により、神官文字の研究は、この 100 年間ほど、大きな進展もなく、立ち遅れた状態にある。実際、現在から 100 年以上も前に執筆された[文献①]が今日でも基本書となっているほどである。[文献①]は古王国時代から末期王朝時代までの字形の変化を通時的に示した良書であるが、現代からみれば、文字判読の精度、文字と形態素の対応、文字の分類方法、文字番号のシステム、聖刻文字との対応関係、コーパスの不足、などの点で修正が必要となっている。加えて、[文献①]では神官文字の字形が手書きで示されているのであるが、そこに記されている字形を原資料で確認すると、時に字形の写しに誤りが認められるのである。

エジプト学におけるこのような事態を改めるためには、聖刻文字翻字至上主義を明確に否定し、原資料の実際の字形を示した[文献①]に代わる神官文字のリストを作成する必要がある。そこで本研究では、神官文字の画像を利用した字形ならびに言語情報のデータベースを作成することにした。

２．研究の目的

本研究は、言語学における文献言語の取り

扱い、ならびにエジプト学における神官文字の取り扱いの問題点を受け、それらの解決を図るべく、(1) 文字情報と語情報の結合、(2) 原資料の写真と言語データの結合、(3) 動的ドキュメンテーションの作成、(4) データ共有、の４つを目標としつつ、データベースの作成を行った。

３．研究の方法

研究の目的で述べた４つの目標のうち、ここでは作業の前提となる目的(1)-(2)の方法について述べる。目的(3)-(4)については「４．成果」で述べることにする。

目的(1)文字情報と語情報の結合

通常、言語テキストは、転写・グロス・訳から構成されるが、このような言語テキストは、語あるいは形態素がすでに設定された状態で示されている。しかし、語や形態素の認定は自明なものではなく、研究者が文字あるいは「文字らしき表記」と対峙することにより、はじめて設定されるものである。そもそも、語や形態素を認定する以前に、その前提として、文字の判読という行為が必要となる。したがって、文献言語学の立場から言うと、語や形態素ばかりではなく、文字に関する情報も言語記述の対象に含まれる。そこで本データベースでは、語・形態素情報（転写・グロス・訳原資料）に加え、文字情報についても記述を行うことにした。

具体的には、写本に見られるすべての文字を読み解き、ページごとに「3-1」のような文字 ID を振った。「3-1」は「3 行目の１番目の文字」を意味する。そして、文字 ID ごとに、文字の属性として「文字コード、インクの色、音、機能、語中の位置」を与えた。神官文字は、コンピューターに入力することのできない“外字”であるため、本研究では Möller 式[文献①]の文字番号を拡張させたコードを独自に設定した。以上が文字情報である。また語については、それぞれの語ごとに転写、品詞、文法範疇、訳を記述した。文字情報と語情報を統合したものを言語データとして扱い、MySQL に格納した。

目的(2)原資料の写真と言語データの結合

目的(1)で作成した言語データに加え、本研究では原資料の写真を撮影し、データベースに格納させた。本データベースで使用した資料は、大英博物館に所蔵されている「神官文字写本 BM10221」である。この写本はラメセス９世時代の墓泥棒裁判の結果を記録したパピルス文書（紀元前 1100 年頃）であり、エジプト学で有名な写本であるにもかかわらず、高精細なカラー写真が公開されていない。そこで、代表者が大英博物館にて資料の写真撮影を行った。使用機材には、中判カメラ Mamiya AFDIII、デジタルバック Phase One P45、Mamiya Sekor Macro MF 120mm f/4 を用いて、約 3900 万画素の高精細画像を得た。

４．研究成果

本研究の実質的な成果は「２．目的」でかかげた目的(3)-(4)である。この２点について以下に述べる。

目的(3)動的ドキュメンテーションの作成

本研究では、原資料の写真と言語データとを格納したデータベースを作成した。BM10221 の未公開カラー写真と言語テキストを格納したという点で、本研究は文献資料のドキュメンテーションとしての価値を有するものとなっている。しかし、本データベースの利点はそのような静的な情報提示にあるわけではない。むしろ、静的なデータを基盤としつつも、そこからデータを動的に生成させる点に特色を持つ 。

次に、本データベースの基本内容について解説する。図１は本データベースのメイン画面であり、上部メニューの Gallery から Page3 を表示させたものである。

この画面の下部に拡大・縮小バーがあり、

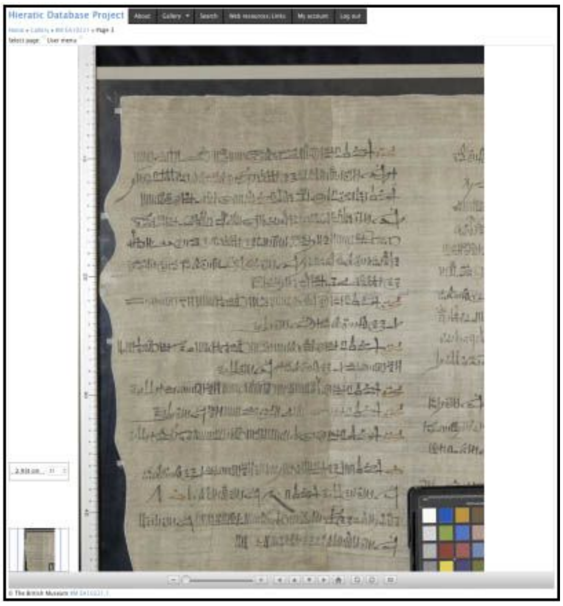

図１

これを使用して写真を拡大させ、文字の細部を確認することができる。また、本ページで任意の文字の上にカーソルを合わせると文字の範囲が自動的に選択され、右側に文字情報と言語情報が示される（図２）。

図２に表示されている文字情報は、文字コード（Code)、文字の音価（Phone）あるいは機能（Function)、備考（Note)、Möller の文字リスト（Möller）である。ここで Möller をクリックすると、Möller の文字リスト[文献①]に掲載されている字形情報がポップアップ画面で示される。また、単語情報では、語の実現形（Realized form)、基本形（Basic form)、文法範疇（Lexical category)、意味(Meaning)、備考（Note)、ならびに語を構成する文字が示される。

語を構成する文字の箇所で、右側にあるサ

ーチマークをクリックすると、その文字が使用されているすべての語が別の画面で表示される（図３）。つまり、文字情報で検索を実行すると、画像を含めた言語情報が生成される。このような動的データの生成が本システムの１つの利点であり、これにより文字や語の記述が容易になる。

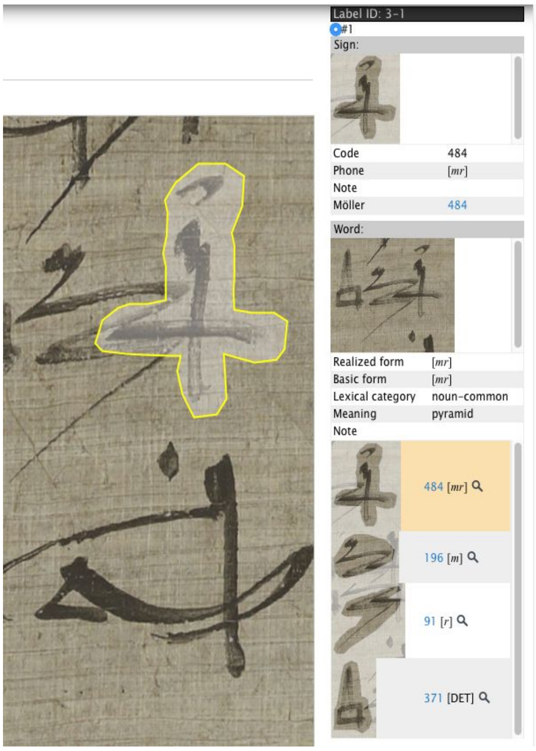

図２

目的(4)データ共有
　動的なものであれ、静的なものであれ、データベースに格納された情報は、そのデータベースでのみ使用可能な閉じた体系となることが多い。しかしながら、言語記述の結果を広く公開し、他の言語データとの比較検討を行うのであれば、データ共有が可能なシステムを構築する必要がある。そこで本システムは、画面に表示されている情報を２種類の XML データで出力する機能を実装させた。１つは文字資料をデジタル化するための国際標準である Text Encoding Initiative (TEI) 形式のデータ出力である。TEI 形式をサポートしたアプリケーションは年々増えており、それらにデータを読み込ませることで、利用者の目的に応じた分析や検索を行うことが可能となる。もう１つは近年提唱されているオープン・データのための RDF (Resource Description Framework) 形式での出力である。ある語の文法範疇が「動詞」であるとき、別のオープンな言語データの「動詞」を自動的に発見することができれば、より大きなデータベース、あるいは「ビッグデータ」を実現することができるが、RDF 形式の出力情報は、「ビッグデータ」に貢献し、検索エンジンによって発見されやすいデータとなっている。

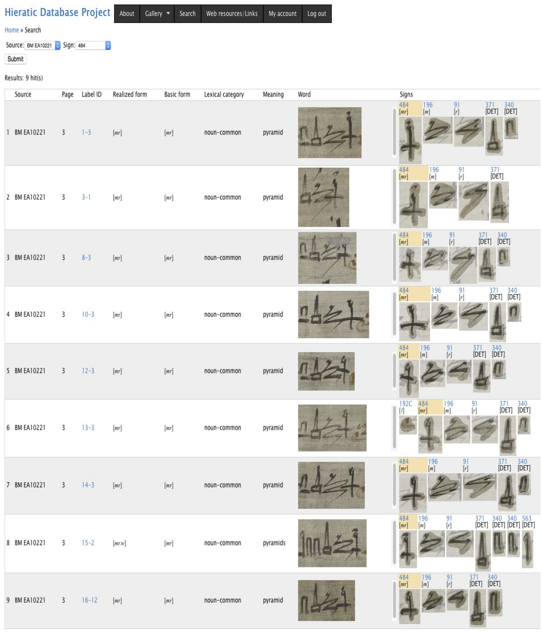

図３

今後の構想
　画像を利用した神官文字のデータベース化は世界初の試みであり、幸いにして海外の研究者から注目を得ている。しかしながら、まだ資料数が少なく、検索システムも完璧ではない。そこで今後はこれらの問題点を克服しつつ、システムを発展させていく所存である。
　実のところ、本データベースで構築したシステムは他言語・他文字に拡張させることができるものであり、実際、現在は研究分担者の和氣ならびに髙橋が中心となり、明治期の日本語文典ならびに楔形文字の粘土板文書を同一のプラットフォームにて作動させるシステムを構築中である。本研究の成果は、古代エジプト語学や言語学ばかりでなく、文字資料を必要とするすべての研究者、ならびに文字資料に関心のある世界の一般ユーザーに貴重な情報を提供するものと期待される。

＜引用文献＞
文献①　Möller, Gorge (1909-1912) *Hieratische Paläographie*. 3 Vols., Leipzig.

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１０件）
[1]髙橋洋成・永井正勝・和氣愛仁「画像、TEI、LOD を用いた文字研究・言語研究のためのプラットフォームの構築」情報処理学会『人文科学とコンピュータ研究会報告』2015-CH-105(5),pp.1-5, 2015 年，査読無.
[2]永井正勝「トリノ・エジプト博物館所蔵の Papyrus Turin Ca.1885 の神官文字に関する覚え書き」『一般言語学論叢』第 16 号, pp.153-169, 筑波一般言語学研究会, 2014, 査読有.

[3]永井正勝「｢モスクワ・パピルス No.120」における接尾代名詞=w の表記とその環境」『一般言語学論叢』第 15 号, pp.41-71, 筑波一般言語学研究会, 2013, 査読有.
[4]和氣愛仁・永井正勝「RDB と CMS を用いたアノテーション付与型画像データベースシステムの構築－データ構造とインターフェイスの標準化を目指して－」情報処理学会『情報処理学会研究報告（人文科学とコンピュータ研究会)』2013-CH-99(7), pp.1-8, 2013, 査読無.
[5]髙橋洋成「アマルナ文書の電子化：文字研究・言語研究を目指して」人文科学とコンピュータ研究会報告 2013-CH-99(6), pp.1-7, 2013, 査読無.
[6]髙橋洋成「言語の多面性を織り込んだ言語資料のデジタルネットワーク」情報処理学会『人文科学とコンピュータシンポジウム論文集』Vol.2013, No.4, pp.39-44, 2013, 査読有.
[7]髙橋洋成「言語資料の電子化に関する諸問題：特に用語法に関して」***Studies in Ethiopian Languages***, Vol.2, pp.49-58, Japan Association of Ethiopian Languages, 2013 年, 査読有.
[8]和氣愛仁・宇陀則彦・永崎研宣・松村敦「閉じる研究と開く研究の接点を目指して－筑波人文情報学研究会の挑戦－」情報処理学会『情報処理学会研究報告（人文科学とコンピュータ研究会)』2013-CH-99(5), pp.1-4, 2013, 査読無.
[9]永井正勝・和氣愛仁「古代エジプト神官文字写本を対象とした言語情報表示システムの試作」情報処理学会『人文科学とコンピュータシンポジウム論文集』Vol.2012, No.9, pp.225-230, 2012 年, 査読有.
[10]Masakatsu Nagai“Some Bibliographical and Graphemic Notes on the Egyptian Hieratic Papyrus 10682 in the British Museum”『文藝言語研究』(言語篇)62, pp.53-74, 筑波大学人文社会科学研究科文芸・言語専攻, 2012, 査読有.

〔学会発表〕（計１０件）
[1]Toshihito Waki, Masakatsu Nagai, Yona Takahashi,“On the Developmnet of a Web-based Platform for Egyptian Hieartic papyri, Akkadian Cuneiform Tables, and Japanese Schoolbooks: Toward a New Grammatological and Linguistic Stuedes”, Digital Humanities Worldwide (「スーパーグローバル大学創成支援」採択事業キックオフワークショップ), 2015 年 2 月 15 日, 筑波大学筑波キャンパス（茨城県つくば市）, 査読無.
[2]髙橋洋成・永井正勝・和氣愛仁「画像、TEI、LOD を用いた文字研究・言語研究のためのプラットフォームの構築」第 105 回人文科学とコンピュータ研究会発表会, 大阪国際大学守口キャンパス（大阪府守口市）, 2015 年 1 月 31 日,審査付.
[3]和氣愛仁,「明治期日本語文典の画像データベースの構築と他言語文化資料への展開」第 27 回日本語教育連絡会議, 2014 年 8 月 24 日, バラトシサールソー（ハンガリー）, 査読無.
[4]髙橋洋成「言語の多面性を織り込んだ言語資料のデジタルネットワーク」」情報処理学会人文科学とコンピュータ研究会/じんもんこん 2013, 2013 年 12 月 14 日-15 日, 京都大学百周年時計台記念館（京都市左京区）, 審査付.
[5]永井正勝「古代エジプト神官文字文書のアノテーション付与型データベース-Hieratic Database Project (HDB) の取り組み-」日本オリエント学会第 55 回大会, 京都外国語大学（京都府京都市）, 2013 年 10 月 27 日, 審査付.
[6]和氣愛仁・永井正勝「RDB と CMS を用いたアノテーション付与型画像データベースシステムの構築-データ構造とインターフェイスの標準化を目指して-」第 99 回人文科学とコンピュータ研究会発表会, 筑波大学東京キャンパス（東京都文京区）, 2013 年 8 月 3 日, 審査付.
[7]髙橋洋成「アマルナ文書の電子化：文字研究・言語研究を目指して」第 99 回人文科学とコンピュータ研究会, 2013 年 8 月 3 日, 筑波大学東京キャンパス（東京都文京区）, 査読無.
[8] Masakatsu Nagai, Toshihito Waki, Yona Takahashi, “On a Hieratic Digital Database of Papyrus Abbott: Toward a New Method of Language Description”, Conference: ***Agyptologische Binsen“-Weisheiten II Von Fruh bis Spat: Hieratisch Phasen Formen und Funktionen der altagyptischen Handschrift***. Mainz (Germany), 2013 年 3 月 20 日, 審査無, 招待講演.
[9]永井正勝「文書の持つ情報の何をどのようにデジタル化するのか？-古代エジプト語神官文字の字形研究を例として-」日本オリエント学会第 54 回大会, 東海大学湘南キャンパス（神奈川県平塚市）, 2012 年 11 月 24 日, 審査付.
[10]永井正勝・和氣愛仁「古代エジプト神官文字写本を対象とした言語情報表示システムの試作」情報処理学会人文科学とコンピュータ研究会/じんもんこん 2012, 北海道大学（北海道札幌市）, 2012 年 11 月 17 日-18 日, 審査付.

〔その他〕
ホームページ等
【Hieratic Database Project】
https://wdb.jinsha.tsukuba.ac.jp/hdb/

６．研究組織
(1)研究代表者

永井　正勝（NAGAI, Masakatsu）
筑波大学・人文社会系・助教
研究者番号：70578369

(2)研究分担者
和氣　愛仁（WAKI, Toshihito）
筑波大学・人文社会系・准教授
研究者番号：70361293
（平成 25 年度より研究分担者）

(3) 研究分担者
髙橋　洋成（YONA, Takahashi）
筑波大学・人文社会系・研究員
研究者番号：90647702
（平成 25 年度より研究分担者）