

# 文字資料を対象とするデータベース構築に適した 言語学的記述のあり方について

永井正勝<sup>†1</sup> 和氣愛仁<sup>†2</sup> 高橋洋成<sup>†3</sup>

**概要**：一般言語学的なスタンスで様々な時代や地域の言語を統一的に扱おうとした場合に、どのような言語学的データの整理の仕方が必要なのかという観点は、データベース構築の際のプラクティカルな問題であると同時に、その整理行為そのものが、言語のあり方を記述する記述言語学の一形態としての価値を有する。本発表では、このような問題意識のもと、文字の直線的な羅列のみを見ても言語構造が見え難いような文字資料をも対象としつつ、文字資料が持つ情報の、何を、どのように、整理・構造化して、それらを情報処理に結びつけていくべきなのか、という点について言語学の立場から提案を行う。

**キーワード**：文字，異体字，字素，音符，限定符，配列，倒置，行間，ヒエラティック，ヒエログリフ

## On linguistic description of written texts aims at making suitable constructions for Linguistic database

MASAKATSU NAGAI<sup>†1</sup> TOSHIHITO WAKI<sup>†2</sup> YONA TAKAHASHI<sup>†3</sup>

### 1. はじめに

今日、デジタル・ヒューマニティーズの研究対象は多岐にわたっており、文字資料、音声、絵画、図像、写真、建物・遺構・遺跡、作品・遺物、地形、気象、風習、食料、人体動作など、人間の活動を中心に様々な側面が扱われている。これらの諸資料に優越を付けるわけではないが、デジタル・ヒューマニティーズの対象としていわば老舗的なものの1つとなるが、文字資料であろう。コンピュータに情報を入力するという作業は何らかの記号を入力することであり、その点で、文字という記号の集積から成る文字資料は、デジタル・ヒューマニティーズと相性のよい資料だと言える。実際、トマス・アキナスの著作の用語集のコンピュータ入力デジタル・ヒューマニティーズの嚆矢であることが、それを物語っている[1]。

しかしながら、文字資料の入力は、実はそれほど単純なものではない。というのは、特に、写本（手書き文字資料）の場合には、難読文字、異体字、書き損じ、修正、欠損など、入力の上でしばしば悩ましい対象となる事象が観察されるからである。これらの困難については、およそ多くの文字資料で問題となることでもあり、それゆえ、文字資料の持つ特徴を生かしつつも、コンピュータ入力上の処理を

統一的に整理し、コンピュータが理解可能なアノテーションを付与するために、Text Encoding Initiative（以下 TEI）[2] が用意され、文字資料のテキスト化で TEI が使用されているところである。

このような状況において、筆者達も TEI を活用した文献資料の整理を行っているが、実は、筆者達が研究対象としている文字資料は、文字入力という点で大きな問題を抱えているものである。特に、筆者の1人が専門とする古代エジプト語の文字資料（ヒエログリフ/聖刻文字、ヒエラティック/神官文字）は、文字が直列に配置されるわけではなく、また限定符と呼ばれる音価を持たない文字があるなど、文字入力そのものに大きな困難を伴うものとなっている。また、ヒエログリフにおいては Unicode が整備されているのだが、文字入力における最大の問題点は、Unicode だけでは文字数が不足していることにある。そのため、ヒエラティックは言うに及ばず、ヒエログリフにおいてすら、文字の全体を統一的に扱うことができないという点で、未だ外字で構成されている資料だと言わざるを得ない。

以上のような難しさを踏まえ、筆者達は「アノテーション付与型画像データベースプラットフォーム」（以下 wdb）の開発を進め、TEI との連携方法を模索してきた[a]。wdb は、エジプト文字の一種であるヒエラティックで書かれた

<sup>†1</sup> 東京大学  
The University of Tokyo

<sup>†2</sup> 筑波大学  
University of Tsukuba

<sup>†3</sup> 東京外国語大学  
Tokyo University of Foreign Studies

a) 具体的な例については、参考文献[3]-[6]ならびに第119回人文科学とコンピュータ研究会発表会（CH119）で発表予定の、和氣愛仁、永井正勝、高橋洋成「アノテーション付与型画像データベースプラットフォームのIIIF対応」を参照。

パピルス写本のデータベース化から始まり、今では近代日本語文字和本資料、古代アッカド語楔形文字粘土板を包含するプラットフォームに成長している。この wdb では開発当初から、言語学の立場でどのように文字を扱うべきかを議論し、試行錯誤を重ねてきた。本稿では特にヒエラティックで書かれた資料を題材として、言語学の立場からデータベースを作成する際に、文字資料が持つ属性のどのような点に着目しているのかという点を整理することにしたい。

## 2. 文字を基点とした言語情報の整理

文字資料には文字が表記されているのであるから、文字入力は資料に表記されている文字を順番に入力すればよいはずであり、たとえば活字で表記された現代日本語の文字資料をコンピュータに入力する場合には、書かれている文字を1つ1つ順に入力すれば済むという点で、大きな困難を覚えることはない。ところが、図1のような古代エジプトのヒエラティック文書の場合はどうであろうか。

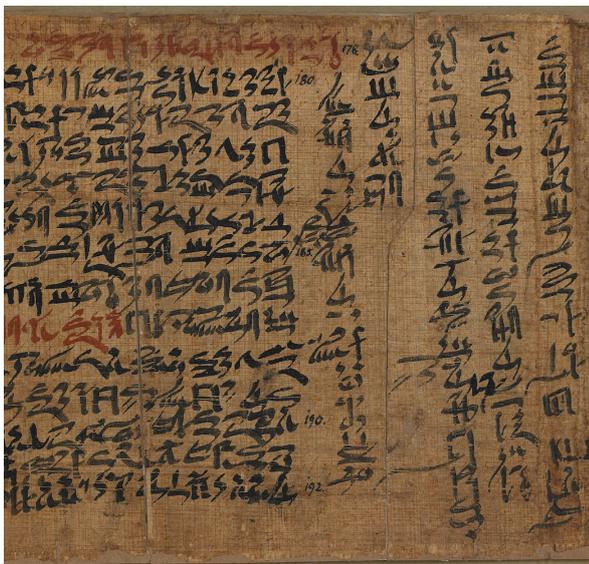


図1 「シヌへの物語」のヒエラティック文書、  
Papyrus Berlin P.3022, 174-192 行目  
Figure 1 The Hieratic text of the “tale of Sinuhe”,  
Papyrus Berlin P.3022, ll.174-192.

図1は古代エジプトの中王国時代(紀元前1800年頃)にヒエラティックで書写されたパピルス写本(Papyrus Berlin P.3022)の一部である[b]。この箇所には、縦書きと縦書きの混在、異体字、表記上の倒置、墨書と朱書き、行間への追記等が見られる。また、ヒエラティック表記一般の特徴として、文字の段組も多用されている。このような資料を扱い、外字として存在している文字をコンピュータに入力し、

言語学に必要な情報を得るためには、少なくとも次のような属性に着目する必要がある。

### 2.1 書字方向

書字方向には、「横書き」と「縦書き」の別に加え、右から左に向けて書く／改行する「右書き」と、左から右に向けて書く／改行する「左書き」とがある。もちろん、より複雑な書字方向もあるが、右縦書き、右横書き、左縦書き、左横書きが主な書字方向となる[8]。

通常、縦書きの文書はすべて縦書き、横書きの文書はすべて横書きとなることが多い。ところが、書字体系によっては縦書きと横書きが混在することがあり、図1はまさにその例である。図1では、図の右方向から、右縦書きが4行続いている(174-177行目)。その後、図中央の上方から右横書きが朱書きで1行だけ書かれている(178行目)。続く179行目は横書きではなく、縦書きとなっている(179行目)。最後に、朱書きされた178行目の下から横書きの文が13行書かれている(180-192行)。つまり、当該箇所は縦4行-横1行-縦1行-横13行というフォーメーションで書かれている。この箇所は、特に複雑な書字方向を持つという点で極端な事例に属するものではあるが、書字方向が混在する資料が存在していることは、データベース設計の際に踏まえておく必要がある。

書字方向に関して、エジプト文字の特徴について言えば、ヒエラティックが右書き専用の文字であるのに対して、ヒエログリフは右書きに加え、左書きも許容される文字である。その際、右書きの文字は動物などの文字の顔が右を向くのに対して、左書きの文字の場合には文字の顔が左を向く。従って、ヒエログリフの場合、右書きと左書きとで文字が鏡文字となる。これが「文字の向き」に関する情報である。上に述べたように、通常、右書きは右向き、左書きは左向きとなるが、文字の向きが逆になるテキストもエジプトの資料には存在する。

### 2.2 文字の配列

右縦書き、右横書き、左縦書き、左横書きなどの書字方向が整えば、あとはその方向に従って、文字を配列することが期待される。ところが、ヒエログリフやヒエラティックでは、文字が直列に配列されず、いわば、行内で段組が行われるのが一般的である。

一例を示す。図1の176行目(右から3行目)に見られる1語を抜き出したものが図2aとなる。

ここには3つの文字が記されており、最初の文字は縦長の文字となり、その左隣に横長の文字が上下二段で配列されている。これをヒエログリフに翻刻すると図2bとなる。176行目は全体として縦書きであるが、1と2(あるいは2+3)を対比させると、横方向に文字が配列されていることになる。

b) 図1-図6に掲載されているパピルスの画像は文献[7]に添付されているDVDより利用したものである。

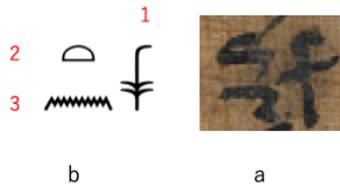


図 2 段組されたヒエラティックの例 (図 1 の 175 行目)  
 Figure 2 An example of hieratic signs arranged in a square  
 (from 1.175 in Fig.1).

もちろん、文字が直列に配される例もあるのだが、行内で段組が行われ、方塊升に収まるように文字が配列されるのが、ヒエラティックとヒエログリフの表記上の特徴である。

このような文字の配列を表示する簡便な入力システムがエジプト学で既に提唱されている。それが“Manuel de Codage” [9]と呼ばれる入力規則である。[9]では、横方向のセットを\*で示し、縦方向のセットを:で示す。それに従えば、図 2b の 1-2-3 の文字配列は、1\*2:3 と表示される。これは、2 と 3 が上下組のセットとなり、このセットと 1 が横方向に並んでいることを示すものである。

### 2.3 文字番号と翻刻

ヒエラティックのような崩し字の場合には、個々の学者が文字の判読を行い、その結果を明示する必要がある。欧米の学者はヒエラティックをヒエログリフに翻字することにより、自らの文字解釈を示すことが多い。しかしながら、ヒエラティックとヒエログリフは十全な意味で 1 対 1 に対応するものではないため、ヒエラティックの翻字としてヒエログリフグリフを使用することは、厳密な意味で翻字とは言えない [c]。

筆者らは、ヒエラティックの翻字として、[11]で採用されているヒエラティックの文字番号を使用することにしており、それに従えば、図 2 にある 3 つの文字列は 289-575-331 で示されることになる。更に、[9]を使用してヒエラティックの文字番号を示すと、289\*575:331 となる。

### 2.4 異体字

ヒエラティックの文字は、多くの場合、異体字を持つ。1 つの文字に対する異体字の数は、個々の学者の判断に依存することではあるが、多くて 3 つ程度だと思われる。図 1 の内部には、たとえば 174 行目と 175 行目にヒエラティック No.33 の異体字が見られる (図 3)。つまり、ヒエラティック No.33 には、ヒエログリフ (図 3c) に近い字体を持つ図 3a と、省略された字体の図 3b とが見られる。



図 3 ヒエラティック No.33 の異体字の例  
 Figure 3 An Example of Allograph of Hieratic No.33.

異体字の弁別は、33a, 33b などのように小文字のアルファベットを付して示すことができる。だが、包括的な文字集合の中から、どれを基本形とし、どれをそのバリエーションとするのかは、通時的ならびに共時的な分析を経て行う必要がある。

### 2.5 文字の色

ヒエラティックはインクを使用したペンで書かれる。インクはおもに黒色 (墨書) であるが、朱色で書かれる部分 (朱書き) もある。朱書きされる箇所は、段落の冒頭や修正を施した箇所であることが多いが、かならずしも規則的に見られるわけではなく、書記によって偏りが大きい。文字情報に色の区別が必要となるという点を理解しておく必要があるだろう。

### 2.6 文字の機能分類: 音符と限定符 (意符)

文字に関する個別の認定が終わると、次は語の識別をしつつ、文字の機能分類を行うこととなる。ヒエラティックとヒエログリフの文字は、表音機能を持つ文字 (音符) と表音機能を持たず、語の意味に関わる文字 (意符) とに大別される。この区別は、漢字の形声を彷彿とさせる。形声の原理で作られた漢字の「江」は、右側の旁 (工) が /ko:/ という音を示し、左側の偏 (氵) が「水」に関する概念を示す。漢字の場合、音符と意符がそれぞれ字素となり、1 字を形成している。それに対して、ヒエラティックとヒエログリフの場合には、音符と意符がそれぞれ独立した文字となる。

図 4a は図 1 の 176 行目に表記されている *hk3* 「支配者」の語を抜き出したものである。この語は 4 つの文字から構成されている。これらの文字が持つ音価をエジプト語研究で使用されている記号で示すと、1 番目の文字は *hk3* (便宜的に /heka/ と読む)、2 番目の文字は *k* (便宜的に /k/ と読む)、3 番目の文字は *3* (便宜的に /a/ と読む) となり、全体として *hk3* と転写され、/heqa/ (へカ) と読まれる [d]。語の意味は「支配者」(男性単数名詞) である。最後の 4 文字目は意符として機能している文字であり、語の意味範疇を示している。「支配者」は人間であるため、ここでは男性の文字、い

c) 翻字 (transliteration) とは、ある文字体系の表記を別の文字・記号体系に置き換えたものを言う。その際、1 対 1 の対応が求められる。詳細は文献 [10] を参照。

d) ヒエログリフとヒエラティックは子音のみを示す文字体系である。本

来は母音を持たないが、エジプト語研究では便宜的に、子音間に /e/ を挿入したり、ある種の子音字を母音として読んだりして、発音しやすいつ形を作成・使用している。

わば人偏と言えるもの、が添えられている。このような意味範疇を示す文字をエジプト語研究では *determinative* 「限定符」と呼ぶ。

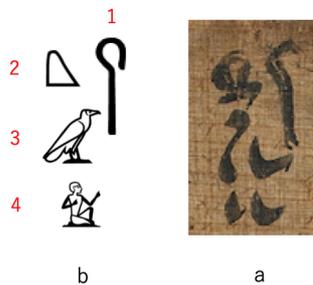


図 4 *hk3* 「支配者」の語の表記例  
Figure 4 Spelling of the word *hk3* “ruler”.

音符として機能している 3 つの文字の音価を順に示すと *heka-k-a* であり、音価の総和は *hekaka* (ヘカカ) となることである。しかしながら、2 番目と 3 番目の文字は 1 番目の文字のいわば振り仮名として機能しているものである。現代日本語の場合、振り仮名は対象となる文字の外側に小さな文字で書かれるが、ヒエラティックやヒエログリフでは、振り仮名として機能する文字が通常の文字と同列に表記される。従って、エジプト文字では、音符の部分において、語の音形を示す文字と振り仮名として機能する文字とが存在することになる。しかも、語の音形を示す文字には、1 文字で 1 音を示す 1 子音文字、1 文字で 2 音を示す 2 子音文字、1 文字で 3 音を示す 3 子音文字が中心となっており、1 文字が 1 音となるわけではない。

語や形態素を決定する上で、このような文字の機能分類の見極めが重要となるが、注意すべき点として、ある特定の文字が、表記上の位置によって、音符となったり限定符となったりする点である。ヒエラティックやヒエログリフでは、文字種と機能が完全に一致しているわけではないため、原資料を読みながら、それぞれの文字の機能を解釈していくことになる。

## 2.7 語の認定

文字の機能の解釈は、その上位の作業として、語の認定を行うことに繋がっており、語の認定結果は語釈（語に関する種々の情報）表示することとなる。図 4 に示した語の語釈は次のようになる。

- 語の音形（子音転写）：*HqA*
- 品詞：名詞（一般）
- 性：男性
- 数：単数
- 意味：支配者

このような語釈を作成することにより、特に文法に関す

る言語分析が容易となる。

## 2.8 文字・語の表記上の倒置

図 1 の 179 行目にある *hpr-k3-r* 「ケペルカーラー（アメンエムハト王の即位名）」という語（王名）を抜き出したものが図 5 である。



図 5 *hpr-k3-r* 「ケペルカーラー」の語の綴り  
Figure 5 Spelling of the word *hpr-k3-r* “Kheper-ka-ra”.

図 5 にある *hpr-k3-r* という語で、*r* の部分が最後に読まれるものであるが、表記を見ると最初に書かれている。その理由は「尊敬の倒置」にある。ヒエラティックやヒエログリフでは、「神名」や神を意味する「*ntr*（ネテエル）」の語に敬意を払い、それらを語頭あるいは語句の冒頭に配置する。これが尊敬の倒置である。図 5 では 神名の *r* 「太陽神ラー」が語頭（あるいは語句の冒頭）に倒置されている。

一方、尊敬の倒置とは関係なく、字形の制約による表記上の倒置もしばしば見られる。図 2 に記されているのは *n:sw:t* 「上エジプト王」を示す語であり、*n-sw-t* という子音から構成されている。これを図 2 に示した数字と対応させると、1=*sw*, 2=*t*, 3=*n* となり、3-1-2 の順で読まれることとなる。その理由は、1 の文字が縦長であるため、1 を含む *sw-t* の部分 (1+2) を倒置させて前に出したものだと思われる。

図 2 ならびに図 5 のような表記上の倒置は、表記されている順番と読む順番とが異なるため、言語データの記述において、二種類の文字配列を用意しておく必要がある。

## 2.9 改行位置

ヒエラティックとヒエログリフでは、1 文字で 1 語となる場合もあるが、おおむね、語は文字列（2 文字以上）で構成される。このように文字列によって語が表記される文字体系を「綴り字型文字体系」と呼ぶことにする。

ヒエラティックやヒエログリフでは、書字材料の長さに応じて、適当な位置で改行を施すことになるが、綴り字型文字体系を採用しているため、改行位置が語境界になるとは限らない。

現代の日本語も綴り字型文字体系に分類されるが、改行位置については機械的に行われることが多く、特に原稿用紙を使用する際や、1 行の文字数が定められている電子ファイルにおいては、行末で機械的に改行されるのが通常で

ある。これはつまり、改行に際して、語の単位へのこだわりが低いことを示す。それに対して、同じ綴り字型文字体系に属する文字体系でも、たとえば英語の表記法では、行末の語にハイフネーションやカーニングを行うことにより、言語単位上の境界で改行が施されるように工夫されている。

ヒエラティックやヒエログリフでも、語中改行がなるべく避けられていたことがうかがえる。図 1 では、174 行目から 179 行目の部分で行末が示されており、これらのすべてが語境界で改行されている。とは言え、書字材料の幅が物理的に制限されているため、語中改行を余儀なくされている行も、本パピルス別の箇所では確認される。改行位置は書記が認知していたと思われる言語単位を知る手がかりとなるため、言語記述の重要な対象となる。

### 2.10 行間への追記

ヒエラティックの資料では、行間に書き込みが見られることがある。図 1 でも、175 行目と 176 行目の間 (図 7a)、および 176 行の左側 (図 7b) に追記された語がある。

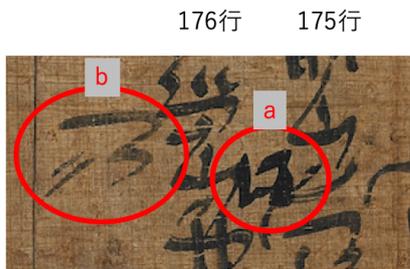


図 6 行間に追記された語の例

Figure 6 Examples of the words added between the lines.

これらが追記だとわかるのは、文脈上、図 7a と図 7b を補う必要があるのと、インクの濃さにある。図 7a は、右隣にある 175 行目の文字よりも濃いインクで書かれている。また図 7b は、右隣にある 176 目の文字よりもインクが薄くなっている。表記していた際に、横にはみ出ただけであれば、インクの濃さが変わることはないであろう。

このように、書記は、筆記を終えた後に推敲を行い、不足した箇所を補ったり、あるいは行間に修正を書き込んだりすることがある。

### 3. 文字という単位認定の難しさ

2 節では、文字という単位を自明なものとして議論をおこなってきた。しかし、実際には、文字単位の認定は実に難しい問題である。

図 7 はヒエログリフの字体とその文字番号[12]を示したものである。ヒエログリフの D4 と U1 はそれぞれ異なる文字であり、D4 の子音転写は *ir*、U1 の子音転写は *m3* である。字形と子音転写の異なる両者は、当然ながら異なる文字として認定される。ところが、両者が連続して表記された U3 は、D4 でも U1 でもない別の文字 (合字) としてエ

ジプト学では認知されており、その子音転写は *m3* となる。U1 と U3 は子音転写が同じであるものの使用される環境が異なるため、別の文字として認定されている。



図 7 ヒエログリフの“合字”の例

Figure 7 An Example of the “ligature” of the Hieroglyphs.

しかしながら、U3 がはたして 1 文字 (合字) であるのか、あるいは 2 文字として認定されるべきなのかは、データベース作成の上で文字種の正規化や文字数のカウントに関わる事項でもあり、悩ましい問題だと言える。U3 が 1 文字だとすると、そこに含まれる D4 と U1 は字素ということになるし、U3 という文字を認めないのであれば、D3 の部分は D4 と U1 が連続した表記として扱われることになる。

次の図 8 の表記も文字単位の認識として悩ましい事例である。

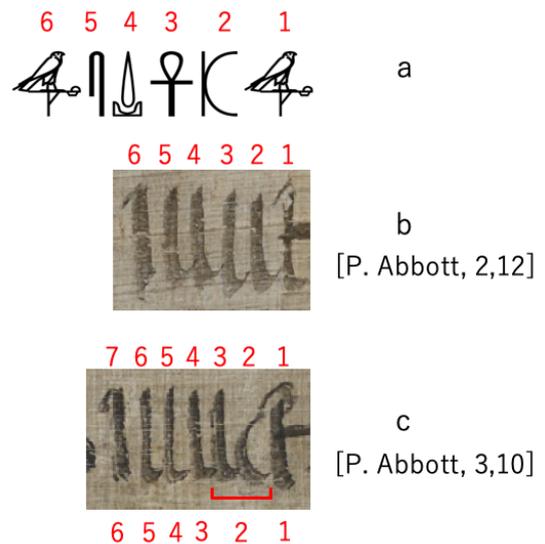


図 8 ヒエラティックの表記例

Figure 8 An Example of Hieratic Writing.

図 8a は王名の後に付加される部分であり、1 が王名に対する限定符、2 が王名を閉じるカルトウーシュ (楕円棒) である。3~5 はそれぞれ *nh(w)* 「行きよ」、*wd3(w)* 「繁榮せよ」、*snb(w)* 「健康であれ」という動詞 (状態形 3 人称男性単数形) となる。最後の 6 は動詞句に付加された限定符となる。

図 8a に対応するヒエラティックには、図 8b あるいは図 8c のような表記が見られる。図 8b の 1 と 6 は文字の上部

に斜線を伴うため弁別的な特徴を持っているが、2~5は同一の字形で表現されており、ヒエログリフの字形と対応させることが難しい。

図 8c では、1 の次に半弧が書かれている。これを 1 文字と見なすのか、それとも文字中の字素とみなすかの判断は、実に難しい。これを 1 文字と見なすと、その後には 5 本の線が書かれていることになる（図 8c の上方の番号の 3~7）。7 は字形の上で 1 と類似しているため、図 8a の 1 および 6 あるいは図 8b の 1 および 6 と同じものだと言える。だが、そうすると、3~6 の部分で、1 文字が余分なものとなる。一方、半弧を字素と見なすと、その次に表記されている縦棒とセットで 1 文字となる（図 8c の下方の番号）。この場合、3~6 は図 8a ならびに図 8b の 3~6 と整合性がつく。その一方で、2 の部分に異体字があるという判断となる。

図 8c の文字解釈については、未だ定まった見解があるわけではないが、どちらの解釈も等しく採用できるようなデータベース設計ができると、人文学として有用なものとなる。つまり、文字らしきものを単位として設定し、登録しておく必要がある。

## 4. データベースへの連結

### 4.1 文字資料を解釈し、入力するということ

文字を入力するとともに、そこに文字や語の属性、言語構造、本文校訂情報等を書き込むことのできる国際的なタグセットとして有名なものが TEI である。実際、TEI の使用は、文字資料の入力に際して、様々な恩恵を我々に与えてくれる。

とりわけ、TEI-P5 (2007 年) において、外字に関する扱いや、原資料画像とテキストとをリンクさせるスタンドオフ・マークアップの方法が定められ、テキスト化しにくい文字資料を扱う方法も整備されつつある。だがなお、ヒエラティックのようにすべてが外字であるような文字資料を扱うまでは至っていない。だとすれば、ヒエラティックの TEI 入力は取り組み甲斐のある試みである。

第 2 節で取り上げた事例を通して、筆者達は以下のことを主張したい。(a)文字そのものに関する情報、(b)文字と文字の関係性、(c)具体的な文脈における文字の機能、これらはデータとして区別されるべきである。ある文献において、どのような文字が現れ、どれを異体字と見なすかは、(a)に属する情報である。書字方向や、文字配列（段組、倒置）といった(b)に属する情報は、次に読むべき文字がどこにあるかに関わる。そして音符として読むか限定符と見なすかといった(c)に属する情報は、語の認定に関わるものである。

このように、文字資料を扱うにあたっては、文字学的情報、文字論的情報、言語学的情報とを区別・整理することが重要である。実際には、文字というものの範囲が難しく、語の定義も難しいというケースが多々ある。だからこそ、それぞれの分野、あるいは「層」における情報を混同する

ことなく、かつ、解釈の多様性を許容するようなデータベース設計であることが理想的である。

### 4.2 データベースと言語学との関係

筆者達が構築している wdb は、開発初期から「辞書の文字集合と文献で使われる文字集合との対応」「ある文字が担いうる機能」「ある文脈の中での文字解釈」、そして「文字句と語との対応」とを区別して扱えるよう設計されている。すなわち、文字学的な層、文字論的な層、言語学的な層を区別しながら、各層の多様な相互関係を表現できるようになっている。そのために、wdb はリレーショナル・データベース（以下 RDB）を基盤として選択した。RDB は集合、関係、階層など、言語学において多用される概念を比較的スムーズに扱うことができる。たとえば、文脈的に離れた位置にある要素同士が関係を結んだり（呼应）、複数の要素に対して複数の要素が結びついたりするなど（多対多の関係）、木構造では表現しにくいような関係を比較的容易に扱うことが RDB では可能である。

また、第 3 節で述べたような、「文字」という単位の認定の難しさについても、RDB はある程度柔軟な対応が可能である。ある記号（wdb ではあえて sign と呼んでいる）が合字を構成する字素であるのか、あるいは独立した文字であるのか、判定に迷うような場合でも、それを「字素」と呼ぶか「文字」と呼ぶかの名付けの問題は別として、ある記号が別の記号とまとまりをなし、さらにそのまとまりが別の記号とまとまりをなすというような再帰的なループ構造を RDB では比較的スムーズに記述できる。すなわち、文字の結合が直ちに語を構成すると決めつけなくても、何層に渡ってでも階層構造を記述できるのである。これはエジプト文字に限らず、漢字でも同じようなことが発生しうるであろう。

もちろん、具体的な実装の手段を選択することが、他の手段を否定することにはならないことは強調しておきたい。ただ、こうした RDB の特徴は、著者達がデータを作成する際に大きな助けになった。たとえば、データを作成するには枠組みの設定が必要となるが、作成作業を通じて判明した言語事実を基に、枠組み自体を修正・更新することもしばしば起こる。このとき、枠組みを比較的柔軟に更新できるということは、RDB による大きなメリットであった。さらに、このようにして修正された枠組みは、単にデータベースに最適化されたということのみならず、言語学的な枠組みの再構築にもつながる。つまり、言語事実をデータベースに適用可能なかたちで分析・整理していく行為そのものが、言語記述の一形態をなしているのである。

「データベースで言語資料を利用する」ということを越えて、「データベースを利用して言語を記述する」への転換は、デジタル技術を利用した言語研究に直接的な意義をもたらす。言語学的な価値を持ちうると、筆者達は考えている。

## 5. おわりに

本稿では、ヒエラティックを題材に文字の持つ属性の整理を行った。繰り返しになるが、筆者達は、ヒエラティックで書かれたパピルス写本の他に、古代アッカド語の楔形文字粘土板と近代日本語文字和本資料をも包含したデータベースを作成している。ヒエラティックと類似して、楔形文字も複雑な属性を持つ文字体系である。「はじめに」で述べたように文字資料はデジタル・ヒューマニティーズと相性の良い資料だと言える。だが、エジプト文字や楔形文字のように、文字そのものを直接的に入力することが難しい文字体系もあるし、そればかりか、文字の解釈そのものが確定していないため、そもそも、どこまでを1文字とするのかという判断が学者によって揺れ動く資料すら存在する。筆者達は、言語情報を扱うデータベースを作成する上で、本稿で概観したような属性に着目し、それに対応しながら、データベースの設計を行ってきた。今後も、文字の持つ属性を分析しつつ、文字資料が持つ多様な情報を扱うことができるように工夫を凝らしていく所存である。

**謝辞** 本研究は JSPS 科学研究費 18K00525 の助成を受けたものである。

## 参考文献

- [1] Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., and J. Schnapp. *Digital\_Humanities*. Mit Press, 2012, p.123.
- [2] “P5: Guidelines for Electronic Text Encoding and Interchange”. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>, (参照 2019-01-24).
- [3] 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. 人文科学とコンピュータシンポジウム論文集, 2012, Vol. 2012, p.225-230.
- [4] 和氣愛仁. 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—. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, 2013, Vol. 2013-CH-99, No. 7, p.1-8.
- [5] 高橋洋成. 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. 人文科学とコンピュータシンポジウム論文集, 2013, Vol. 2013, p.39-44.
- [6] 高橋洋成, 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. 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, 2015, Vol. 2015-CH-105, No. 5, p.1-5.
- [7] Parkinson, R., *Four 12th Dynasty Literary Papyri (Pap. Berlin P.30322-5): A Photographic Record*. Akademie Verlag, 2012.
- [8] 永井正勝. 古代エジプト聖刻文字の書字方向—一般統字論構築の一助として—. 一般言語学論叢, 2005, 第 8 号, p.21-45.
- [9] “Manuel de Codage”. <http://www.catchpenny.org/codage/>, (参照 2019-01-24).
- [10] Coulmas, F. *The Blackwell Encyclopedia of Writing Systems*. Blackwell, 1999, p.511-512.
- [11] Möller, G. *Hieratische Paläographie*. 3vols.. J.C. Hinrich's Buchhandlung, 1909-36.
- [12] Gardiner, A. *Egyptian Grammar: Being an Introduction to the Study of Hieroglyphs*. 3rd ed., Griffith Institute, 1957, p.442-543.