

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 2 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26870085

研究課題名(和文) 前14世紀の楔形文字文書、アマルナ書簡の言語記述のためのデジタルアーカイブ構築

研究課題名(英文) Digital Archiving of the Amarna Letters: Towards the Linguistic Study of the Cuneiform Tablets in the Fourteenth Century B.C.E.

研究代表者

高橋 洋成 (TAKAHASHI, Yona)

筑波大学・人文社会系・非常勤研究員

研究者番号：90647702

交付決定額(研究期間全体)：(直接経費) 1,100,000円

研究成果の概要(和文)：本研究は前14世紀にパレスチナからエジプトに送られた外交文書(アマルナ書簡)のデジタル・アーカイブを構築するものである。アマルナ書簡の特徴は、第一に楔形文字粘土板であること、第二に当時の国際語であるアッカド語と現地語であるカナン語との混成言語で書かれていることである。そこで、各粘土板の写真を画像データベースに登録するとともに、楔形文字情報ならびに言語分析情報とを連携させ、データの一体化を図った。これにより本研究は多層的なデータをひとまとめに扱うためのデジタル・プラットフォーム研究と、言語の総体的な記録を目指す言語ドキュメンテーション研究の両面にわたる基礎と実例を提供するものとなった。

研究成果の概要(英文)：The aim of this research is to construct a digital archive of the Amarna Letters, which were the diplomatic letters mainly from Ancient Palestine to Egypt in the fourteenth century B.C.E. These letters are important, not only in that they were inscribed by the cuneiform scripts on the clay tablets, but also in that they were written in the mixed language of both the official Akkadian and the local Canaanite language. This research succeeded to combine three different types of data in one digital platform, such as photos of the clay tablets stored in the image database, each cuneiform information and the linguistic analysis on the mixed language of the letters, by using the XML tagset provided by Text Encoding Initiative (TEI). The research provides the theoretical and practical basis for the future study on the Digital Platform, as well as for the philological and general linguistic studies.

研究分野：言語学

キーワード：アマルナ書簡 デジタルアーカイブズ 楔形文字 アッカド語 カナン語 文献言語学 Text Encoding Initiative Linked Open Data

1. 研究開始当初の背景

(1) 近年、デジタル記録装置の大容量化・高性能化にともない、人文学資料のデジタル化についても新たな方法論が求められ、議論が続けられている。言語学においては、言語記述を目指す従来の言語コーパス作成から、言語の使用状況まで含めた総体的な記録(言語ドキュメンテーション)が提唱されている。すなわち、動画や音声、写真といった言語実態の記録と、言語分析に関する情報とを一体化させようという試みがなされている。

(2) 楔形文字資料のデータベース化については、これまでも世界的に多くのプロジェクトが進められている。しかし、データベースは各研究者の研究目的によって、データの精度や取捨選択の方法、分類方法、使用方法が異なっている。そのため、それぞれのデータベースを連携させてより「大きなデータベース」にしていくことが困難であった。

2. 研究の目的

(1) 楔形文字資料の中でも、アッカド語とカナン語の混成言語で書かれているアマルナ書簡(前14世紀パレスチナからエジプトへ送られた外交書簡)のデジタル・アーカイブを作成する。この過程において、楔形文字情報のデータベース化ならびに共有化、混成言語であるアマルナ語の文法記述と、デジタル技術を用いた文法範疇の共有化を行い、文献言語資料のデジタル化について一般性・相互運用性の高い方法論の確立を目指す。

(2) (1)で作成された楔形文字データと言語分析データに加え、画像データベースと連携することにより、写真画像・文字・言語といった多層的な情報を扱う「文献言語ドキュメンテーション」のための方法論を確立する。同時にその運用方法と、それともなう新たな研究分野の可能性について、研究者との議論を深めていく。

3. 研究の方法

(1) アマルナ書簡の楔形文字をローマ字転写したテキストデータはすでに公開されているが、デジタル的に構造化されたものではない。そこで、人文学資料のデジタル化をリードする国際グループ Text Encoding Initiative (TEI)が提供しているXMLタグセットを用いてテキストを構造化する。具体的には、楔形文字の符号化ならびに楔形文字辞典とのリンク、語および語構成に関する文法情報のマークアップを行う。

(2) 時間的な制約上、約380通のアマルナ書簡のうち研究期間内にデジタル化するものを限定しなければならない。そこで、アマルナ語の混成言語としての特徴が強く現れているもの30通を選択する。

(3) TEI/XMLでマークアップされた文字情報・言語情報をオープンデータとし、楔形文字粘土板の画像と連携できるようにする。特に、研究協力者である筑波大学の和氣愛仁准教授が開発しているアノテーション付与型画像データベースシステムと連携し、粘土板画像・文字情報・言語情報を1つのプラットフォームで取り扱うことを目指す。

4. 研究成果

(1) 研究期間内に作成されたアマルナ書簡データは次の通り(なお、アマルナ書簡はEAの通し番号で管理される)。楔形文字情報の符号化ならびに楔形文字辞典とのリンクについては、298通(EA60~339、362~378、381)の粘土板についてデータを作成した。

文法情報のマークアップについては、Amurruを発信地とする5通(EA60~64)、Gubla(Byblos)を発信地とする5通(EA73、77、82、86、87)、Megiddoを発信地とする5通(EA242~246)、Urušalim (Jerusalem)を発信地とする6通(EA285~290)、Gezerを発信地とする9通(EA267~271、292~295)、合計30通の書簡データを作成した。

(2) 19世紀に楔形文字が解読されて以来、粘土板を研究資料として公開・共有するための一定の形式が確立されている。たとえば、次に挙げるのはEA61粘土板の裏面7行目(EA61 rev:7)をローマ字転写したものである。

[.....]LUGA]L d.UTU EN-i[a

楔形文字の読み方には漢字と同じように音読み(シュメル語読み)と訓読み(アッカド語読み)が存在するが、ローマ字転写は原則として音読みで記述する。もしある文字が音節文字的に読まれるものであれば小文字にし、それらが語中に並ぶ場合はハイフンでつなく。もし音節文字的ではなく表語文字的に読まれるものであれば大文字にし、解釈する際に文脈に応じて適切な訓読みを当てる。また、漢熟語のように表語文字的なものが2つ以上並んで1つの訓読みを持つときはピリオドでつなく。このとき語の意味クラスを表す文字(漢字の部首に相当)があるなら小文字にする。さらに、上記の例では粘土板の破損箇所が[...]のように示されている。また、LUGA]Lは文字の左側が不明瞭であること、i[aは文字の右側が不明瞭であることを示す。このように、楔形文字粘土板のローマ字転写は粘土板の状態、文字の解釈、語の範囲といった多層的な情報を、研究書という紙媒体が持つ制約のゆえに、ひとまとめに扱うというものであった。それは一方ではあたかもアスキーアートのように粘土板を視覚的に再現しつつ、他方では言語的解釈をも示すという、研究者の努力の結晶であった。だが、本研究は紙媒体の制約に左右されないデジタルデータを作成するものである以上、

これまでひとまとめに扱われていた情報の層を切り分けねばならない。すなわち、粘土板画像、文字解釈データ、言語解釈データを独立したものとして作成し、かつそれらを互いにリンクさせるような仕組みを考える必要がある。以下に、その主要な成果を挙げる。

(3) シュメル・アッカド系の楔形文字は、2000年に発足した Initiative for Cuneiform Encoding (ICE)によって Unicode 化が進められ、2004年に第1段階が完了した。その成果が Unicode 5.0 に収録された U+12000 区画に属する 879 字、および U+12400 区画の数字・句読点 103 字である。ICE は第1段階においてウル第3王朝(前22~21世紀)のシュメル字体を採録しており、今後は字源をさかのぼって古アッカド体や古代の字体を採録していく方針となっている。

しかしながら、本研究が対象とするアマルナ書簡のほとんどは中期バビロニア字体(前16~10世紀)で書かれており、字形のみならず字体も Unicode に対応していないものが少なくない。たとえば、名詞の複数性を表す MEŠ は頻出する文字であるが、Unicode にしたがえば ME. EŠ のように字源をさかのぼった2字として表現しなければならない。

そこで、本研究は Unicode に加えて3つの楔形文字字典、すなわち Borger(Borger 1978)、Syllabar (von Soden 1991)、MesZL (Borger 2003) における楔形文字番号および文字の読み方を整理し、TEI/XML タグセットを用いてデジタルデータ化することで、楔形文字同士の関連性を示せるようにした。

この際、2つの可能性が考えられた。第1の可能性は、TEI/XML の辞書モジュールを使用し、各楔形文字字典を辞書としてマークアップする方法である。しかし、この方法は文字の読みに関するデータの大部分が重複するという欠点を抱えていた。そこで第2の可能性として、本研究では TEI/XML の外字モジュールを使用し、楔形文字を1字ずつ外字として定義するという方法を採用した。ただし、外字モジュールだけでは文字の読みに関する情報、たとえば楔形文字の音読みと訓読みに関する情報を十分に記述することができなかった。そこで、独自の拡張要素として reading 要素を設け、下記のように value 要素の列挙によって文字の読みを示すようにした。さらに value 要素には言語タグを付与し、音読み(シュメル語読み、@xml:lang="sux")と、訓読み(アッカド語読み、@xml:lang="akk")、あるいは数字(@xml:langなし)であることを明示した。

```
<char>
  <charName>AŠ</charName>
  <charProp>
    <localName>Unicode</localName>
    <value>ASH</value>
```

```
    <value>U+12038</value>
  </charProp>
</charProp>
  <localName>MesZL</localName>
  <value>001</value>
</charProp>
<charProp>
  <localName>Borger</localName>
  <value>001</value>
</charProp>
<charProp>
  <localName>Syllabar</localName>
  <value>001</value>
</charProp>
<reading>
  <value>1</value>
  <value lang="sux">AŠ</value>
  <value lang="sux">DIL</value>
  <value lang="sux">DILI</value>
  <value lang="akk">ina</value>
  <value lang="akk">aplu</value>
  <value lang="akk">maga: ru</value>
  <value lang="akk">nada: nu</value>
  <value lang="akk"
cert="medium">se: pu</value>
  <value lang="akk">we: du</value>
  <note>cf. MesZL: U. DAR (670+183)</note>
</reading>
<reading>
  <value lang="sux">U2. AŠ</value>
  <value lang="akk"
cert="medium">e: du</value>
</reading>
</char>
```

このように機械処理可能な楔形文字データは世界的に見ても未だ公開されておらず、今後はこれをオープンデータとして研究者の間で共有し、フィードバックを頂いてより精緻なものにしていく方針である。

(4) 世界の言語の多様性を反映し、文法情報をデジタル化する方法も多種多様である。それに加え、電子コーパスに文法情報を埋め込む場合、その研究目的・制作コストを勘案した範囲での作業になるため、コーパスごとに文法情報をデジタル化する方法が異なるという状況も生じている。

そこで、言語コーパスのデジタル化に際し、必要となりうる語彙をウェブ・オントロジーとして管理しようとするのが GOLD、ISOcat、OLiA などのプロジェクトである。また、オープンデータの立場からは Linguistic Linked Open Data (LLOD) が提唱され、コーパス・辞書・文法情報などを広く共有していこうという世界的な動向がある。詳細は高橋(2013b)を参照されたい。

本研究はアマルナ語の語形成分析を進めながら、一般性と記述しやすさを両立した文法記述の確立を目指した。アマルナ語はアッカ

ド語とカナン語との混成言語であり、語形成にも両方の言語の性質が表れている。たとえば、 \emptyset -iprus「私は切った」と iprus- \emptyset 「彼が切り」という、見た目は同一の形姿であるにも関わらず、異なる屈折範疇に属する2つの語は、アッカド語とカナン語の両者の文法的特徴が入り混じって生じたものである。

このような形姿と意味との不一致を扱えるよう、本研究は次のように文法情報を記述した。まず土台として用いるのは TEI/XML タグセットである。この中にはコーパス作成モジュールや簡単な言語分析モジュールも用意されているが、実際にどう使用するかはデータ作成者に委ねられている部分が多い。本研究は、語を表す w 要素、形態素を表す m 要素、正規化要素を表す reg 要素を採用し、さらに独自の拡張要素として言語特徴を示す feature 要素を導入した。

たとえば、yisallaluši「彼は彼女(町)を略奪している」(EA292:45)のマークアップを考えてみたい。アマルナ語はセム諸語の特徴を受け継ぎ、不連続形態素や象徴法(内部変容)がよく用いられる。そこで形態素境界を示すことを保留し、語が持ちうる意味的特徴を列挙したのが以下である。

```
<w>
<reg>yisallaluši</reg>
<m>
  <feature type="class">verb</feature>
  <feature type="pos">verbal</feature>
  <feature type="theme">G</feature>
  <feature
type="stem">imperfective</feature>
  <feature type="person">third</feature>
  <feature
type="gender">masculine</feature>
  <feature
type="number">singular</feature>
  <feature
type="object
person">third</feature>
  <feature
type="object
gender">feminine</feature>
  <feature
type="object
number">singular</feature>
</m>
</w>
```

このように語を「かたち」よりもまずは意味的特徴の束と見なす形態意味論的分析)については、Anderson(1992)が提唱する A-Morphous Morphology を参照されたい。だが繰り返し述べるように、本研究は混成言語であるアマルナ語の形姿と意味との不一致に着目するものである。したがって、上述した意味的特徴それぞれに、必要に応じて形態音韻論的な「かたち」の情報を埋め込む必要がある。

<w>

```
<reg>yisallaluši</reg>
<m>
  <feature type="class">verb</feature>
  <feature type="pos">verbal</feature>
  <feature
type="theme"
pattern="prs">G</feature>
  <feature
type="stem"
pattern="iparras">active</feature>
  <feature
type="tma"
suffix="u">imperfective</feature>
  <feature
type="person"
prefix="y">third</feature>
  <feature
type="gender"
prefix="y">masculine</feature>
  <feature
type="number"
prefix="y">singular</feature>
  <feature
type="object
person"
suffix="si">third</feature>
  <feature
type="object
gender"
suffix="si">feminine</feature>
  <feature
type="object
number"
suffix="si">singular</feature>
</m>
</w>
```

上記では、形態法として接辞 (@prefix、@suffix)、内部変容のパターン (@pattern) が示されている。このように意味と形態とを結び付けておけば、たとえば「語幹 (@type="stem") が能動的 ("active") である feature 要素を持ち、そのかたちが iparras パターン (@pattern="iparras") であるもの」のような組み合わせ検索が可能になる。また、上記の例は下記のように、m 要素でグループ化することもできる。どちらの書き方がよいかはデータ作成者の好みの問題であるが、後者の方が、語幹と接辞という形態の切れ目が鮮明になるという特徴を持つ。

```
<w>
<reg>yisallaluši</reg>
<m>
  <feature type="class">verb</feature>
  <feature type="pos">verbal</feature>
  <feature
type="theme"
pattern="prs">G</feature>
  <feature
type="stem"
pattern="iparras">active</feature>
  <feature
type="tma"
suffix="u">imperfective</feature>
</m>
<m prefix="y">
  <feature type="person">third</feature>
  <feature
type="gender">masculine</feature>
  <feature
type="number">singular</feature>
</m>
<m type="object" suffix="si">
  <feature type="person">third</feature>
```

```

<feature
type="gender">feminine</feature>
<feature
type="number">singular</feature>
</m>
</w>

```

以上、TEI/XML の既存のタグセットを土台とし、ただ1つの独自拡張要素として feature 要素を設けることで、形態意味論の観点からの記述と、形態音韻論的な情報とを両立させ、複雑な形態法であっても低コストで記述できることが分かった。このように TEI/XML 用いた文法記述の詳細についてはウェブサイトを参照されたい。

(5) 楔形文字データと文法項目データとを、研究協力者である筑波大学の和氣愛仁准教授が開発しているアノテーション付与型画像データベースシステムと連携することを目指し、そのプロトタイプを構築した。図1は、そのシステム上で楔形文字の範囲を選択してみた例である。

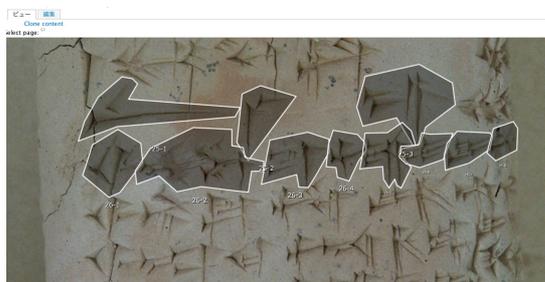


図1 楔形文字を範囲選択する

以上のように、楔形文字粘土板の写真画像と、文字情報、および言語情報という多層的なデータ群を、1つのプラットフォームで可視化することに成功した。

研究資料のデジタル化は、単にオンラインで資料を閲覧することやデータベース化することを越えて、様々な資料を一体化させるためのデジタル・プラットフォームを提供するものへとつながる。このように言語現象を総合的に処理し、統一的なプラットフォーム上で様々な観点から処理してみることは、これまで見過ごされてきた事象をすくい上げ、新たな学術領域を開拓していくことにもつながる。本研究は楔形文字資料を言語学資料として扱うという見地から、そのための基礎理論と、実際の運用例を提供するものである。

<引用文献>

Andersen, S. R., Cambridge University Press, A-Morphous Morphology, 1992, 452

Borger, R., Neukirchener Verlag, Assyrisch-babylonische Zeichenliste, 1978, ix+452

Borger, R., Ugarit-Verlag, Mesopotamisches Zeichenlexikon, 2003, viii+736

Soden, W. von, Röllig, W., Pontificio Istituto Biblico, Das akkadische Syllabar, 1991, xli+76

高橋 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、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-CH-99(6)、2013、1-7

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 1件)

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、査読無、人文科学とコンピュータ研究会報告 2014-CH-105(5)、2014、1-5

[学会発表](計 3件)

TAKAHASHI, Yona, A Digital Archive of the Amarna Tablets: Diplomatic Relations and Their Lingua Franca in Ancient Palestine, 14th century B.C.E., Eighth Forum for the Euro-Japan Academic Networking for Humanities Project: Trajectories from Pre-Modern to Post-Modern, November 7-8th 2015, Ljubljana (Slovenia)

永井 正勝、和氣 愛仁、高橋 洋成、古代エジプト語神官文字資料のコーパス作成とその利用：“外字”で書かれた文字言語のドキュメンテーション化の試み、日本言語学会第150回大会、2015年6月21日、大東文化大学板橋キャンパス(東京都板橋区)

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、第105回人文科学とコンピュータ研究会、2015年1月31日、大阪国際大学守口キャンパス(大阪府守口市)

[その他]

ホームページ等

<https://wdb.jinsha.tsukuba.ac.jp/xsux/>

6. 研究組織

(1)研究代表者

高橋 洋成 (TAKAHASHI, Yona)

筑波大学・人文社会系・非常勤研究員

研究者番号：90647702

(2)研究分担者 なし

(3)連携研究者 なし

(4)研究協力者

池田 潤 (IKEDA, Jun)
筑波大学・人文社会系・教授
研究者番号：60288850

和氣 愛仁 (WAKI, Toshihito)
筑波大学・人文社会系・准教授
研究者番号：70361293

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