

Building Web Corpus of Old Nubian with Interlinear Glossing as Digital Cultural Heritage for Modern-Day Nubians

So Miyagawa and Vincent W.J. van Gerven Oei

Kyoto University, Punctum Books & Coventry University

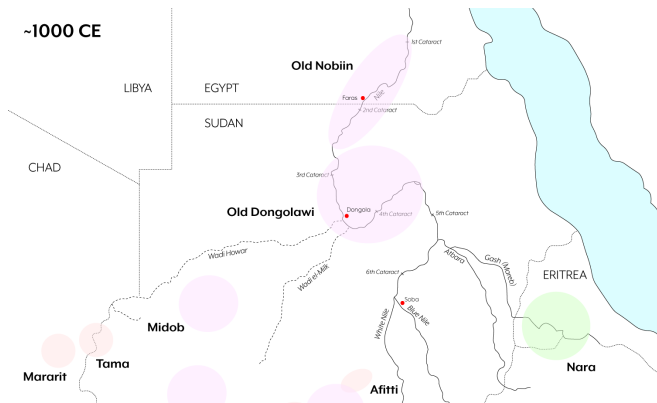
September 8, 2021

Contents

- 1 Old Nubian
- 2 Goal of the project
- 3 Corpus-building methodology
- 4 Pilot project
- 5 Future perspectives for digital cultural heritage

The Old Nubian language

- A Nubian language
- Recorded in the kingdoms of Nobadia and Makuria in the Middle Nile Valley
 - Modern-day southern Egypt and northern Sudan
 - Between the 8th and 15th centuries CE



The Nubian alphabet

Second oldest in Nilo-Saharan

Besides Meroitic, Old Nubian is the oldest written language from the Nilo-Saharan language phylum.

- Written using the Coptic-derived Nubian alphabet
- Coptic alphabet: Greek alphabet + 6-7 Demotic-derived letters
- Includes three characters from the Meroitic alphasyllabary

ϥⲟⲓⲗⲁ ⲃⲟⲩⲣⲉ· ⲁⲡⲟⲓⲗⲓⲕⲁ ⲃⲁⲓⲛⲓⲁ·

Old Nubian texts

- Old Nubian corpus consists of both 1. literary material and 2. documentary materials
 - 1 Such as Bible translations and sermons, and burial texts
 - 2 Such as contracts, land sales, wall inscriptions



Project goal

Problem so far...

There is no digital text corpus of Old Nubian.

Team for the Digital Old Nubian Corpus project

So Miyagawa digital humanist and Coptologist

Vincent van Gerven Oei Old Nubian language expert

Our goal: linguistically and philologically tagged corpus

- User-friendly, highly visual online digital platform
- Data sustainability and interoperability through TEI XML

Linguistic tagging for Old Nubian texts

- Visualization for both linguistic experts and Nubian heritage holders
 - 1 For linguists, interlinear glosses following Leipzig Glossing Rules (LGR)
 - Glosses under each word and morpheme
 - 2 For Modern-Day Nubians, Arabic and plain English grammatical annotations
 - Nubians: heritage holders of Old Nubian



Corpus-building methodology

Corpus data originally composed in X₃LA_TE_X

Interlinear glossing realized in four lines by gb4e.sty

- 1 Old Nubian text in the Old Nubian alphabet
- 2 Romanized version of the same Old Nubian text
 - Morphemes parsed by hyphens
- 3 Interlinear glossing
- 4 English translation

```

360 \begin{exe}
361 \exlabel{x:M-115-8}
362 \textsf{M} 1.5--8
363 {{\ON{\No\Nstr\Nt\No\Nu}}{\ON{\No\Nu\Ne\NN\No}}}{\ON{\Nd\Np\Nstr\Np\No\Nu}}{\ON{\No\Nu\Ne\NN\Na}}
{\ON{\Nd\No\Nu\Na\Nstr\Nr\Na}\textperiodcentered}{\ON{\Na\Nstr\NN\Ne\Nz\Na\Nn\Nd\Nr\Ne\Nn}\textperiodcentered}
{\ON{\Nsh\Nk\Nstr\Ng\No\Nu\NNa}\textperiodcentered}
364
365 \gll eitrou ouel-$\varnothing$-lo dippou ouel-la dou-ar-a alexandre-n \v{s}ik-gou-la \l
366 woman one-\textsc{nom-foc} village one-\textsc{dat} exist-\textsc{pst1-pred} Alexandria-\textsc{gen} district-ground-\textsc{dat}\l
367 \trans There was a woman living in a village, in the district of Alexandria.'
368 \end{exe}

```


The entire Old Nubian corpus

- The entire Old Nubian corpus is not as big as that of Coptic and far smaller than that of Ancient Greek
- Large amounts of epigraphic materials
- The most sizable is Pseudo-Chrysostomus's *In venerabilem crucem sermo*: 3,200-word
- The second most sizable text is *the Miracle of Saint Mina*: 950 words.

■ Pls: Caroline T. Schroeder & Amir Zeldes



Help-Basics | **Query Result**

Base text: <K> << 1 / 1 >> Displaying Results 1 - 1 of 1

Left context: 5 | Right context: 5

Path: Basics > > an_ark_of_food_norm_group B1 - 46

norm_group norm_group_pos lemma orig_group orig lang translation

norm	norm	pos	lemma	orig_group	orig	lang	translation						
norm	norm	pos	lemma	orig_group	orig	lang	translation						
pos	ADV	ADV	PREP	PPRO	N	PUNCT	CONEJ	ACCUS	PPER	NHET	V	PPER	N
lemma	norm	pos	lemma	orig_group	orig	lang	translation						
orig_group	norm	pos	lemma	orig_group	orig	lang	translation						
orig	norm	pos	lemma	orig_group	orig	lang	translation						
lang	norm	pos	lemma	orig_group	orig	lang	translation						

Hence, let us all pray, everyone who fears God, that Christ may put his blessing end that it may remain in his own regions a place's lack anything.

p_n 1.2
pb_xmli_id BB553
cb_n 1
lb_n 24 25 26 27 28
tok norm pos lemma orig_group orig lang translation

analysis: word [document]
diplomatic text [document]
normalized text [document]

Starting point

Pilot project of digital Old Nubian corpus

- Stauros Text: up to 800 words, relatively long
- Vincent van Gerven Oei developed the interlinear glossed text file in XeLaTeX
- So Miyagawa created XSLT to convert this XeLaTeX file into TEI XML format

```
<ab xml:id="SC4">
  <s type="orig">ⲉⲁⲉⲓⲁ ⲟⲩⲕ ⲟⲕⲓⲟⲩⲁⲡⲡⲉ</s>
  <s type="parse">ⲉⲁⲉⲓ-ⲁ ⲟⲩ-ⲕ ⲟⲕ-ⲓⲟ-ⲁⲡ-ⲡ-ⲉ</s>
  <s type="roman">ⲙⲁⲉⲓ-ⲁ ⲟⲩ-ⲕ ⲟⲕ-ⲓⲙ-ⲁⲣ-ⲣ-ⲉ</s>
  <s type="gloss">who-QUOT 2PL-ACC call-PLACT-INTEN-PRS-1SG.PRED</s>
  <s type="trans" xml:lang="en">'What shall I call you?'</s>
  <note>Notes The following affirmative forms in -ⲙⲁ are all dependent on the verb
    ⲟⲕⲓⲟⲩⲁⲡⲡⲉ.</note>
</ab>
```

Making it into a webpage

- XSLT program to transform the TEI XML into an HTML file
- Combined with a JavaScript file (leipzig.js by Benjamin Chauvett)
- Enables visualization of our interlinear glosses in the form of LGR
- This pipeline produced the first online Old Nubian corpus

Γαῑῑδ̄ ογκ̄ ὀκῑδ̄αρρε̄

Γαῑῑδ̄

ογκ̄

ὀκῑδ̄αρρε̄

η̄aei-a

ou-k

ok-ij-ar-r-e

who-QUOT

2PL-ACC

call-PLACT-INTEN-PRS-1SG-PRED

'What shall I call you?'^[2]

- More Old Nubian literary texts with LGR-styled interlinear glossing via TEI
- To facilitate an online search, we will create a search function of each lemmata and glosses on the homepage using XQuery
- Providing photos of Old Nubian manuscripts in IIIF
- First digitized lexicon data of Old Nubian for further NLP development for Old Nubian

13 / 15

Future perspectives for digital cultural heritage 2

- Primarily aimed at Old Nubian philology experts or linguistics, or historians in Medieval Nubia
- But planning to embed
 - plain English and Arabic translations
 - easy-to-read explanations for each gloss
 - a basic grammar of Old Nubian
 - lecture videos
- To contribute to Nubian cultural preservation and heritage education
- To spread knowledge of the Nubian culture to a wider global audience.

Thank you.

For further questions and comments: miyagawa.so.36u@kyoto-u.jp

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers:

- JP20K21975, Grant-in-Aid for Research Activity Start-up, “Studies on Vowel System and Phonetic Values of Doubled Vowel Letters in Old Nubian,”
- JP21K00537, Grant-in-Aid for Scientific Research (C), “Studies of Coptic Vowels in Historical Phonology of Egyptian Language”