

The ‘TraCES’ Project: Towards a New Approach to Studying the Gə‘əz Language

Eugenia Sokolinski, Universität Hamburg

The project TraCES introduces a novel research infrastructure for the computer assisted study of the Gə‘əz language as it is transmitted in written sources from various periods. This newly created corpus of annotated Ethiopic texts opens new horizons to linguists and philologist and can be subjected to an array of scientific queries.

Gə‘əz, frequently referred to as (Classical) Ethiopic, a Semitic language of the South Semitic branch, offers a particular case study for a Semitic language that developed on African soil, with remarkably little influence from non-Semitic Afro-Asiatic languages (Cushitic). The language can boast extensive documentation in a vast array of written sources going back two millenia. The earliest extensive witnesses are monumental inscriptions, but the overwhelming bulk is formed by literary texts transmitted in manuscripts, the oldest being datable to Late Antiquity and the most recent to the twentieth or even twenty-first century. The texts include early translations from Greek, later translations from Arabic, and local compositions. This situation calls for a diachronic approach to language study, rooted in a thorough morphological and lexical analysis of the literary sources.

To set up the tool box for such analysis has been, since 2014, the task of the project *TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages* at the Hiob Ludolf Centre for Ethiopian Studies of Universität Hamburg.¹

The texts

At the core of the TraCES project is a corpus of literary texts that have been judged representative of a historical period or type of transmission, corresponding thus to a particular layer of Gə‘əz written heritage.²

- 1 This project was funded by the European Research Council (ERC) under the European Union’s Seventh Framework Programme (grant agreement no. 338756, PI: Alessandro Bausi, Universität Hamburg), <<https://www.traces.uni-hamburg.de/>>. See also Bausi 2015a.
- 2 On digitization approach to the texts, see Sokolinski 2016.

For the current project phase, the selected layers have been: (1) Gə'əz of the Aksumite period;³ (2) Gə'əz of the Aksumite translations from Greek;⁴ (3) Gə'əz of mediaeval translation from Arabic;⁵ (4) original texts composed in Gə'əz (even if these may as well be partially inspired by translated models).⁶

Among the original texts, the *Kəbra nagašt* represents a special case of a text outstanding for its peculiar literary development in romance form of a subject well-known in several traditions. Most probably compiled and fixed in this form in the first half of the fourteenth century, with possibly at least partial 'Vorlage' in Copto-Arabic milieu (depending on the degree of trust one places in the 'colophon'), it incorporates elements from various traditions of the Near East (Christian, but also Jewish, with Islamic parallels) for which expression had to be sought in Gə'əz. While historians of Late Antiquity tend to highlight the significance of *Kəbra nagašt* for early relationship and acquaintance with Christian Near Eastern history and literatures,⁷ others have remarked the ambiguity of the textual evidence.⁸ The lexical and grammatical features might be pointing to a complex and varied translation history.

Research prerequisites

As a first step, a research infrastructure had to be created.

An XML database, set up in cooperation with another Hamburg-based project, *Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea*,⁹ hosts the authority files associated with the texts and the named entities occurring in the texts.

The extensive metadata on texts includes structured information on the titles and labels under which the works may be circulating in Ethiopian tradition; on the authorship and attributions; on the existing editions and translations as well as secondary publications discussing the works in question; on the Vorlage, recensions, and versions existing in other languages. Each text is

3 This layer is illustrated by vocalized inscriptions RIÉ 187, RIÉ 188, RIÉ 189, RIÉ 193I, RIÉ 193II, RIÉ 195I, RIÉ 195II, and RIÉ 232. The edition is based on the transcription in the *RIÉ* volumes.

4 This layer is represented by the texts of the Gospel of Matthew, see Zuurmond 2001, the *Testamentum Domini*, see Beylot 1984, and the *Epistle of Eusebius to Carpianus*, see Bausi 2015b.

5 E.g. the *Life of Secundus the Silent*, ed. Heide 2014.

6 Here, three very different text types have been selected, for the moment all from the earlier phase of Ethiopic literature: the national epos, the *Kəbra nagašt* (ed. Bezold 1905), the theological anti-heretic *Book of Mysteries* (ed. Həruya 'Ērməyās 2007), and the historiographic *Chronicle of 'Amda Šəyon* (ed. Marrassini 1993);

7 See e.g. Shahīd 1976; Debié 2010.

8 See e.g. Bausi 2010, 2016; Piovaneli 2013.

9 See the project note by Dorothea Reule in this issue.

assigned a unique ID consisting of the alphanumeric sequence corresponding to the number associated with the work in the *Clavis Aethiopica* (CAe, created primarily by the *Beta maṣāḥaft* project), prefixed by 'LIT' and followed by a short title.¹⁰ Each work description is linked to the detailed description of manuscript testimonies to the work as well as to records dedicated to the author(s), translator(s), commissioner(s), and other persons involved in the creation and transmission of the text.

These latter records are part of the extensive authority files repertory containing data on persons and places relevant for the Ethiopian manuscript culture, including all those mentioned in the texts that are being annotated in the *TraCES* project. In this way, each personal or place name found in the text can be directly linked to the record containing extensive metadata on the person or place in question, including name variants, historical background, geographical data, etc.¹¹

At the core of the research infrastructure is the digital dictionary of Gə'əz.¹² The dictionary had as its base the text of the *Lexicon linguae Aethiopicae* by Dillmann.¹³ The entire text has been retyped and normalized (by Alessandro Bausi and Andreas Ellwardt), and each lemma was (automatically) supplied with a unique ID (by Cristina Vertan). Subsequently, the TXT file was upconverted from string to XML using regex and XSLT (by Pietro Liuzzo). Each lemma is encoded as an extensive TEI-XML file (using TEI Dictionary module);¹⁴ further encoding is possible with the help of a web-based editor. The dictionary is visualized online with the help of an application based on the eXist-db, developed by Pietro Liuzzo. The word list has been significantly enlarged having included lemmata not listed by Dillmann but registered in the texts processed by the *TraCES* project, or in other dictionaries or indices of the Ethiopic language. The dictionary is linked to the text corpus placed online by the *Beta maṣāḥaft* project, meaning that in addition to the examples provided by the lexicographers, all occurrences of the

10 E.g., the CAe number for the *Kəbra nagašt* is 1709; its ID in the database is LIT1709Kebran, the record can be viewed at <<http://betamasaheft.eu/LIT1709Kebran>> with the source code being <<http://betamasaheft.eu/LIT1709Kebran.xml>>. The CAe number for the *Chronicle of 'Amda Šəyon* is 4275; its ID is LIT4275ChronAmdS (<<http://betamasaheft.eu/LIT4275ChronAmdS>>, etc.

11 See e.g. PRS1854Amdase, the record on King 'Amda Šəyon, or LOC6136Waj, the record on the region of Wağ.

12 <<http://betamasaheft.eu/Dillmann/>>, last accessed 20 June 2018.

13 Dillmann 1865. See also the paper by Wolfgang Dickhut and Andreas Ellwardt in this issue.

14 See e.g. the lemma ቡባኤ ፋ, <<http://betamasaheft.eu/Dillmann/lemma/L8f5fa9c212e24d4e99b1a64b7d1e4205.xml>>.

lemma in actual texts can be easily visualized.¹⁵ While currently only the exact forms can be searched for, an engine is being developed by Stefan Druskat to include a morphological parser and a set of grammatical rules that will enable the web applications to simultaneously search for all declined and/or conjugated forms of a certain lemma.

The extensive morphological tag set for annotating Ethiopic, developed within the *TraCES* project framework, is another important prerequisite (and a by-product in its own right). The annotators can select among 33 distinct part-of-speech tags, each supplied, wherever necessary, by a set of additional grammatical features (e.g., for the Verb, the features include Person, Gender, Number, Tense/Aspect/Mood; the Infinitive takes the nominal features of Case and State).¹⁶

Work tools

The project would not be possible without the help of a vast array of electronic tools assisting the researchers.

The main tool is the GeTa annotator, a Java application developed by Cristina Vertan specifically for the *TraCES* project.¹⁷ The tool allows automated and synchronized Latin transcription of Ethiopic texts to be manually corrected and then used as the basis for mark-up at several levels. Transcribed graphic units can be split into minimal meaningful morphological units (tokens); each token is then associated with a part-of-speech tag (with features) and linked to the relevant lemma in the electronic dictionary (fig. 1).

At the same time, the tool allows for a basic philological annotation, including text divisions (chapters, sections, verses, lines, sentences, etc.), line and page breaks, as well as marking up the Named Entities, including persons, places, titles, offices, and dates. Personal names and toponyms are linked to the aforementioned authority files, enabling full indexing and disambiguation of terms. Comments can be introduced at all levels.

While the tool stores the data in JSON format, various converters allow exporting data into an array of other formats, ensuring the sustainability of the project data and interoperability with other projects.

Thus, thanks to the TEI-XML converter, the *TraCES* texts with philological mark up can be visualized on the *Beta maṣāḥəft* web application.

15 For ቡባሊ ፋ, see the visualization at <<http://betamasafeft.eu/Dillmann/lemma/L8f5fa9c212e24d4e99b1a64b7d1e4205>>.

16 See Hummel and Dickhut 2016 for an introduction into the tag set, with a focus on the annotation of nouns.

17 See Vertan 2016 for the preliminary description of the tool functionalities as well as the paper by Susanne Hummel, Vitagrazia Pisani, and Cristina Vertan in this collection.

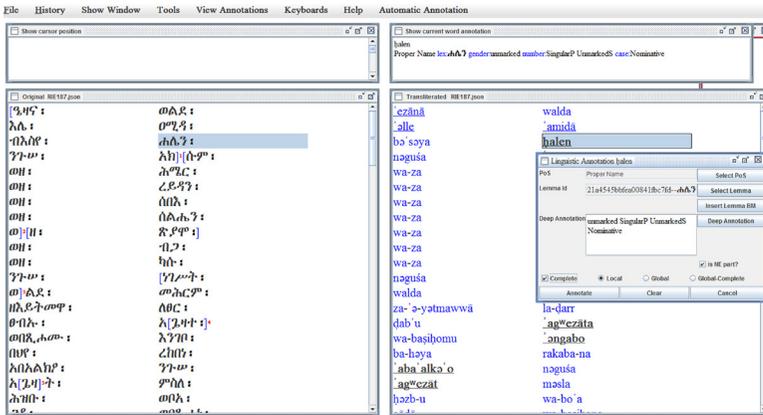


Fig. 1. GeTa application window (annotation of RIÉ 187).

ባህሪ:	ሰነድ:	ወይንስ:	ቶቶ:	ጀምላኝ:	ገዳላ:	ቢረኤኦት:	ላዋዚሲ:	ሊሰሰ:	ክርባኝ:	ገዳ:	ወወዊላ:	ዘገባ:	ጠዘላሳላሳ:	ባህሪ:	ላዋዊ:	ጽገኑ:	ገንጽ:	ሊትጽጽ:	ወይንስ:	
FIDED	ባህሪ:	ሰነድ:	ወይንስ:	ቶቶ:	ጀምላኝ:	ገዳላ:	ቢረኤኦት:	ላዋዚሲ:	ሊሰሰ:	ክርባኝ:	ገዳ:	ወወዊላ:	ዘገባ:	ጠዘላሳላሳ:	ባህሪ:	ላዋዊ:	ጽገኑ:	ገንጽ:	ሊትጽጽ:	ወይንስ:
TR	ba-soma	ab	wa-wald	wa-manfas	qaddus	1-amlak														
grid (anno)																				
FID	ባህሪ	ሰነድ	ወይንስ	ቶቶ	ጀምላኝ	ገዳላ	ቢረኤኦት	ላዋዚሲ	ሊሰሰ	ክርባኝ	ገዳ	ወወዊላ	ዘገባ	ጠዘላሳላሳ	ባህሪ	ላዋዊ	ጽገኑ	ገንጽ	ሊትጽጽ	ወይንስ
FIDEDh	ባህሪ	ሰነድ	ወይንስ	ቶቶ	ጀምላኝ	ገዳላ	ቢረኤኦት	ላዋዚሲ	ሊሰሰ	ክርባኝ	ገዳ	ወወዊላ	ዘገባ	ጠዘላሳላሳ	ባህሪ	ላዋዊ	ጽገኑ	ገንጽ	ሊትጽጽ	ወይንስ
TR	ba-soma	ab	wa-wald	wa-manfas																
FIDLET	ባ	ሰ	ወ	አ	ብ	ወ	ወ	ል	ጅ	ወ	ወ									
FIDLETED	ባ	ሰ	ወ	አ	ብ	ወ	ወ	ል	ጅ	ወ	ወ									
TRFID	ba	sa	ma	a	b	wa	wa	l	d	wa	ma									
TOKL	ba	soma	ab	wa	wald	wa	manfas													
NT	Preposition	Common Noun	Common Noun	Conjunction	Common Noun	Conjunction	Common N													
case		Nominative	Nominative				Nominative													
gender		Unmarked																		
gender-nature			Masculine																	
gender-pattern																				
gender-syntax																				
lex	ባ	ሰ	ወ	አ	ብ	ወ	ወ	ል	ጅ	ወ	ወ									
logogram																				
number																				
number-pattern		Singular	Singular																	
number-syntax																				
person																				
state	Nominal state	Construct state	Absolute state																	
tam																				
ne	false	false	true																	
NRI	1																			
NR	1																			
LE	1																			
NA	sentence																			
CR	Annotator																			
NET																				
R		http://betamasahfeft.eu/api/PRS3697egziabe/tei	http://betamasahfeft.eu/api/PRS5684JesusCh/tei	http://betamasahfeft.eu/api/PRS5684JesusCh/tei	http://betamasahfeft.eu/api/PRS5684JesusCh/tei															
T		person	person	person	person															
Name		ጌጊጊ ጳጳሌ	Jesus Christ	Paráqillos																
Occupation		rel.	Son of God																	
Sex		1	1																	
tok	b	a	s	e	m	a	a	b	w	a	w	a	l	d	w	a	m	a		

Fig. 2. ANNIS search fields for a GeTa-annotated text (here: *Chronicle of Amda Šayon*).

Thanks to the GeTa-to-ANNIS ‘Pepper’ converter, implemented by Stefan Druskat, the linguistic annotation can be now stored, visualized, searched, and analyzed online, with the help of the web-based tool for linguistic mark up, ANNIS (ANNotation of Information Structure; fig. 2).¹⁸ Texts annotated in the *TraCES* project, alongside the relevant metadata, shall be hosted in the sustainable portal of the Hamburger Zentrum für Sprachkorpora (HZSK).¹⁹ Users can then search and analyze texts using the local ANNIS instance maintained by HZSK.

Outlook

With the help of the annotated corpus and the search and visualization tools it shall be possible for the first time to attempt a new diachronic picture of the language which, even if not spoken for most of the past millennium, has remained the only medium of written expression until well into the nineteenth century. For the first time, frequency analysis can be carried out for lexical and grammatical phenomena; changes in linguistic patterns, in word use, and stylistic choices can be first analysed and accounted for on an objective basis.

References

- Bausi, A. 2010. ‘The Massacre of Najrān: The Ethiopic Sources’, in J. Beaucamp, F. Briquel-Chatonnet, and C. J. Robin, eds, *Juifs et chrétiens en Arabie aux V^e et VI^e siècles. Regards croisés sur les sources*. Le massacre de Najrān, II, Centre de recherche, d’histoire et civilisation de Byzance, Monographies, 32 (Paris: Association des Amis du Centre d’Histoire et Civilisation de Byzance, 2010), 241–254.
- 2015a. ‘TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages’, with an annex by C. Vertan, in A. Bausi, A. Gori, D. Nosnitsin, and E. Sokolinski, eds, *Essays in Ethiopian Manuscript Studies. Proceedings of the International Conference Manuscripts and Texts, Languages and Contexts: the Transmission of Knowledge in the Horn of Africa, Hamburg, 17–19 July 2014*, Supplement to *Aethiopica*, 4 (Wiesbaden: Harrassowitz Verlag, 2015), 11–15.
- 2015b. ‘La versione etiopica della Epistola di Eusebio a Carpiano’, in R. Zarzeczny, ed., *Aethiopia Fortitudo ejus: Studi in onore di Monsignor Osvaldo Raineri in occasione del suo 80° compleanno*, *Orientalia Christiana Analecta*, 298 (Roma: Pontificio Istituto Orientale, 2015), 107–135.
- 2016. ‘La leggenda della Regina di Saba nella tradizione etiopica’, in F. Battiatto, D. Hartman, and G. Stabile, eds, *La Regina di Saba: Un mito fra Oriente e*
- 18 <<http://corpus-tools.org/annis/>>. See also Druskat and Vertan 2018.
- 19 <<https://corpora.uni-hamburg.de/>>. The launch of the TraCES corpus online is scheduled for autumn 2018.

- Occidente, Atti del Seminario diretto da Riccardo Contini, Napoli, Università 'L'Orientale', 19 novembre 2009–14 gennaio 2010*, Archivio di Studi Ebraici, 8 (Napoli: Centro di Studi Ebraici, Dipartimento Asia, Africa e Mediterraneo, Università degli Studi di Napoli 'L'Orientale', 2016), 91–162.
- Bausi, A. and E. Sokolinski, eds 2016. *150 Years after Dillmann's Lexicon: Perspectives and Challenges of Gə'əz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016).
- Beylot, R., ed., tr. 1984. *Testamentum Domini éthiopiens. Édition et traduction* (Louvain: Peeters, 1984).
- Bezold, C., ed., tr. 1905. *Kebrā Nagast. Die Herrlichkeit der Könige, nach den Handschriften in Berlin, London, Oxford und Paris zum ersten Mal im Äthiopischen Urtext hrsg. und mit dt. Übers. versehen*, Abhandlungen der Philosophisch-Philologischen Klasse der Königlichen Bayerischen Akademie der Wissenschaften, 23/1 (München: Verlag der K. Akademie der Wissenschaften in Kommission des G. Franz'schen Verlags, 1905).
- Debié, M. 2010. 'Le Kebrā Nagast ou La Gloire des rois: une réponse apocryphe aux événements de Najran?', in J. Beaucamp, F. Briquel-Chatonnet, and C. J. Robin, eds, *Juifs et chrétiens en Arabie aux V^e et VI^e siècles. Regards croisés sur les sources*. Le massacre de Najrân, II, Centre de recherche, d'histoire et civilisation de Byzance, Monographies, 32 (Paris: Association des Amis du Centre d'Histoire et Civilisation de Byzance, 2010), 255–278.
- Dillmann, C.F.A. 1865. *Lexicon linguae Aethiopiae, cum indice Latino. Adiectum est vocabularium Tigre dialecti septentrionalis compilatum a W. Munziger* (Lipsiae: T.O. Weigel, 1865; repr. New York, 1955, and Osnabrück: Biblio Verlag, 1970).
- Druskat, S. and C. Vertan 2018, 'Nachnutzbarmachung von Forschungsdaten und Tools am Beispiel altäthiopischer Korpora', in G. Vogeler, ed., *Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum, 26.02.–02.03.2018 an der Universität zu Köln*, (Köln: Universität zu Köln, 2018), 270–273.
- Heide, M., ed., tr. 2014. *Secundus Taciturnus. Die arabischen, äthiopischen und syrischen Textzeugen einer didaktischen Novelle aus der römischen Kaiserzeit*, Mit einem Beitrag von S. Weninger zur *Übersetzungssprache der äthiopischen Version*, Äthiopistische Forschungen, 81 (Wiesbaden: Harrassowitz Verlag, 2014).
- Ḥərūya 'Ermoḃyās, ed. 2015. መጽሐፈ ፡ ምሥጢር ፡ ዘአባ ፡ ጊዮርጊስ ፡ ዘጋሥጫ ፡ (*Maṣḥafa məšṭir za-³abbā Giyorgis za-Gaśəččā*, 'The Book of Mystery of ³Abba Giyorgis of Gaśəččā') (Addis Ababa: Alem Printing House, 2007).
- Hummel, S. and W. Dickhut 2016. 'A Part of Speech Tag Set for Ancient Ethiopic', in Bausi and Sokolinski 2016, 17–30.

- Marrassini, P., ed., tr. 1993. *Lo scettro e la croce. La campagna di 'Amda Šeyon contro l'Ifāt (1332)*, Studi Africanistici, Serie Etiopica, 4 (Napoli: Università degli Studi di Napoli l' 'Orientale', 1993).
- Piovanelli, P. 2013. 'The Apocryphal Legitimation of a 'Solomonic' Dynasty in the Kəbrä nägäšt— A Reappraisal', *Aethiopica*, 16 (2013), 7–44.
- RIÉ. É. Bernard, A.J. Drewes, and R. Schneider, *Recueil des Inscriptions de l'Éthiopie des périodes pré-axoumite et axoumite*, introduction de Fr. Anfrais, I: *Les documents* (Paris: Diffusion de Boccard, 1991).
- Shahîd, I. 1976. 'The 'Kebra Nagast' in the Light of Recent Research', *Le Muséon*, 89 (1976), 133–178.
- Sokolinski, E. 2016. 'The TraCES Project and Gə'əz Studies', in Bausi and Sokolinski 2016, 13–16.
- Vertan, C. 2016. 'Bringing Gə'əz into the Digital Era: Computational Tools for Processing Classical Ethiopic', in Bausi and Sokolinski 2016, 31–42.
- Zuurmond, R. 2001. *Novum Testamentum Aethiopice*, III: *The Gospel of Matthew*, Aethiopistische Forschungen, 55 (Wiesbaden: Harrassowitz Verlag, 2001).