breeding grounds for disgruntled young men, while at the same time they could not prevent the decline in Buddhist and Pali studies, as evidenced by poor examination results in these areas.

Dhammasami's study breaks new ground by investigating systematically Buddhist monastic learning and its curricula over several centuries and in two Buddhist countries. As the author shows, this topic was not an internal affair of the sangha alone but a crucial element of its complex and sometimes fraught relationship with kings or secular authorities more generally. His investigation into the forms, motives and results of royal interference in monastic learning provides a well-documented addition to the debate concerning the sangha-state affairs, which is usually focussed on the issue of "purity" and "purification" of the sangha and the canon.

**Tilman Frasch**
Manchester Metropolitan University

AFRICA

ALESSANDRO BAUSI (ed.):
*150 Years after Dillmann's Lexicon: Perspectives and Challenges of Gəʿəz Studies.*
(Supplement to Aethiopica 5.) xi, 238 pp. Wiesbaden: Harrassowitz, 2016. ISBN 978 3 447 10783 9.
doi:10.1017/S0041977X19000612

The book under consideration is a collection of eleven papers that were all but one presented at the conference *150 Years after Dillmann's Lexicon: Perspectives and Challenges of Gəʿəz Lexicography* at the University of Hamburg in October 2015, during the initial phase of the research project *TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages.* The TraCES project has as its objective the creation of both an annotated digital corpus of Ethiopic texts and a digital lexicon of Gəʿəz interlinked with the corpus. This is a promising undertaking, considering that digital, annotated resources will not only facilitate the search for specific attestations of words and word forms in texts from different eras, but will also allow for quantitative research of linguistic phenomena.

The core part of the present book consists of a number of papers dedicated to specific aspects of the TraCES project. Eugenia Sokolinski ("The TraCES project and Gəʿəz studies") describes the work plan of the project. One major challenge addressed in this plan is that most Gəʿəz texts are typically not readily available in a digital, Unicode-encoded format; therefore preprocessing is necessary. Printed texts are digitized using OCR (Optical Character Recognition) software and manually post-corrected. Digital texts available in outdated encoding are transformed to Unicode data by use of macros. A text is then automatically indexed and transliterated, after which it is manually tokenized, lemmatized and annotated using the GeTa annotation tool, which was specifically developed for the project.

The GeTa annotation tool is described in more detail by Cristina Vertan ("Bringing Gəʿəz into the digital era"). It has been designed for manual annotation, yet allows semi-automation in the sense that batch annotation of multiple tokens is

possible. Annotation can be done on different levels (graphic unit, token, edition, text structure). The annotation data is stored in JSON format and is compatible with the TEI standard widely applied in the humanities.

As Vertan rightly notes, fine-grained tokenization is only possible in transliteration, because token boundaries may lie inside a syllable (e.g. ቤቱ *bet-u*). The annotation on the token level is thus done solely on the transliteration. Nevertheless, the original text in the Ethiopic script is saved in parallel. A simple but valuable feature of the GeTa tool is that it automatically synchronizes the original text in the Ethiopic script and the transliteration after every modification of either.

The GeTa tool has been developed because apparently no existing tool allows for the complexity of annotations intended for the project. The question therefore arises as to whether it is planned to make the annotation software available to other researchers at some point in the future. This question is not addressed in the paper. In any case, the tool might be of interest for other projects with similar objectives.

Susanne Hummel and Wolfgang Dickhut ("A part of speech tag set for Ancient Ethiopic") introduce the PoS (part of speech) tag set applied in the project. Somewhat unfortunately, the PoS tags are distributed very unevenly across the basic word classes for no obvious reason. Of the 33 tags, only two are dedicated to nouns (common nouns and proper nouns) and there is a single tag for all verbs. Almost half of the tags are reserved to distinguish different types of particles, resulting in very specific tags such as "deictic imperative particle".

A token is annotated not only with a part of speech tag, but additionally with relevant features such as gender, number, case and state for nouns. In the annotation of these features, not just morphology, but also syntax is taken into consideration. If neither morphology nor syntax is decisive on the status of a given feature, the respective feature (e.g. number) is tagged as *unmarked*. A difficult case in question is gender, since the gender system in Gəʿəz is not as stable as it is in other classical Semitic languages. According to Hummel and Dickhut's guidelines, nouns are considered to have a gender "by pattern" only if the same pattern exists with and without a feminine marker. For instance, ሕዝብ *ḥəzb* ("people") is considered to be *unmarked* for gender (p. 24), because it does not stand in contrast to a morphologically feminine counterpart. The rules developed by Hummel and Dickhut seem overly complicated on first impression. At any rate, they are not fully comprehensible from the paper alone. For instance, the token ሙሴ *muse* ("Moses") is annotated as *masculine* "by nature" (as opposed to "by pattern"), yet as *unmarked* for number (p. 25). However, Moses is presumably not only *masculine*, but also *singular* "by nature".

Of the remaining chapters, which are not directly concerned with the TraCES project, the excellent paper by Maria Bulakh ("Some problems of transcribing Geez") deserves special mention. Bulakh notes that romanization of Gəʿəz is typically neither strict transliteration nor strict transcription. While most romanizations tend to be transliterations, they typically include two features not present in the Ethiopic script, viz. consonant gemination, and presence or absence of *ə*. Since these features cannot be extracted from the Ethiopic script, one has to rely on either the traditional pronunciation or a linguistic reconstruction of the morphology. However, these two sources contradict each other in some instances. The paper concludes with a detailed discussion of selected Gəʿəz words in which the presence or absence of *ə* is debatable.

Some other notable papers include Stefan Weninger ("The use of Arabic in Gəʿəz lexicography"), who provides a critical assessment of Arabic etymologies and cognates in August Dillmann's *Lexicon* and Wolf Leslau's *Dictionary*; Alessandro Bausi ("On editing and normalizing Ethiopic texts"), who discusses different

solutions for orthographic normalization in critical editions of Ethiopic texts; and Andreas Ellwardt ("Beyond Dillmann's *Lexicon*"), who presents an interesting historical overview of Gəʿəz and Syriac lexicography. To conclude, the present book provides valuable insights both on the history of Gəʿəz lexicography and on aspects of digital processing of Gəʿəz resources, which will likely also be of interest for scholars of neighbouring disciplines.

**Benjamin Suter**
University of Zurich