Anne Multhoff

9 A Methodological Framework for the Epigraphic South Arabian Lexicography. The Case of the Sabaic Online Dictionary

Abstract: This paper describes the concept and functionalities of an online reference dictionary for Sabaic, aiming to present all extant lexical material of this Ancient South Arabian language. After introducing the features of the corpus, several methodological issues, and the solutions adopted within the project are illustrated, focusing in particular on the annotation of morphological analysis (treatment of ambiguous forms, homographs, heterographs with identical meaning, variant readings, incorrect forms). The conception developed to present the material online is also described.

Keywords: Sabaic, Ancient South Arabian, lexicography, variants, translations

9.1 Introduction

9.1.1 General Remarks

Over recent decades, a huge amount of new Ancient South Arabian inscriptions has come to light. This has been published in collections such as Arbach & Schiettecatte, 2006 or Priolella, 2013 to name but a few, but mostly in scattered editions comprising only a few texts. This material not only contains hitherto unknown lexemes, but also calls for a reconsideration of quite a number of well-known terms in the Sabaic lexicon. The available dictionaries on Sabaic such as Beeston et al., 1982 or Biella, 1982 thus no longer reflect the present state of research. The same holds true for dictionaries on other Ancient South Arabian idioms such as Qatabanic (Ricks, 1989) and Minaic (Arbach, 1993). Moreover, apart from a considerable quantitative increase of the material, we are also confronted with a qualitative leap, as completely new text genres have emerged, particularly among the everyday correspondence on wooden sticks (cf. Stein, 2010 & Maraqtan, 2014). Though revision of at least part of the material included in the extant dictionaries was rightly demanded in both well-meaning (e.g. Lundin, 1987) and cynical reviews (cf. Jamme, 1985, pp. 202–269), a revised second edition of Beeston et al., 1982 never appeared. An up-to-date presentation of the

Anne Multhoff, Friedrich-Schiller-Universität Jena

© 2018 Anne Multhoff

This work is licensed under the Creative Commons Attribution 4.0 International license (CC BY 4.0)

lexical data gathered over the past 30 years is therefore clearly a desideratum of the scientific community, both within and outside Ancient South Arabian Studies proper.

9.1.2 Scope of the Project

The project described in the present paper¹ aims at a reconsideration of the lexical material. It will result in a reference dictionary (“Belegwörterbuch”) that will include a complete lexical survey of the Sabaic material published so far. In contrast to other projects on Ancient Arabian epigraphy featured in this volume, such as DASI and OCIANA, it is not focused on the epigraphic corpus as such, but uses the latter as a basis for lexicographic work. Digitization of material is thus not considered as a result intended for public use, but rather as a practical means to collect and organize large amounts of data.

The application is running on a Microsoft-Windows-Server on which the IIS (Microsoft) is installed as internet server. The data is stored on an instance of the Microsoft SQL Server Express. Applied programming languages are C# and JavaScript. While the internal working platform was designed in ASP.Net, the more modern ASP. Net MVC is used for the public web presence. Furthermore, the JavaScript framework JQuery is used in the web presence.

Two different concepts, adjusted to the various parts of the working process, are used for data management. First, a collection of the epigraphic material is needed as a material base to reference lexemes. For the annotation of texts an XML format was chosen. An editing view of each annotated text is generated from the XML document as a HTML view via a JavaScript routine. Annotations are directly assigned to words as XML attributes. XML documents thus generated are stored in the database. Following the grammatical analysis, the information contained within these XML documents is divided into the tables of the relational database to enable further processing.

Second, a complete set of the interpretations given in the literature is collected. This is to facilitate lexicographic work, comprising translations from text editions, extant dictionaries and further lexicographical material published both in compendia (e.g., Sima, 2000) and specialized articles (e.g., Robin, 2013), or studies on Ancient South Arabian culture and history (e.g., Beeston, 1976) to name but the most important genres. The collected material is further enriched with etymological parallels from other Semitic languages. This material has no intrinsic relation to a specific inscription. It is directly stored in the tables of the relational database.

1 [<http://sabaweb.uni-jena.de>].

9.2 Material Base

9.2.1 Character of Material

Sabaic is part of a group of several interrelated epigraphically documented Semitic languages, commonly referred to as Ancient South Arabian, which were spoken in the territory of modern Yemen from the early 1st millennium BCE up to the 6th century CE (cf. Stein, 2011). The material, however, is rather extensive both in respect to absolute length of attested texts and as far as lexical variance is concerned, at least if compared to contemporaneous European texts of equal genre, as e.g. Latin inscriptions. Reading is considerably facilitated by the regular use of word dividers². As with most Semitic languages, the script is highly defective: basically only consonants are noted, with the only exception of final long vowels.³ Abbreviations are absent. Inscriptions as material objects are written on durable material, mostly stone or metal. They thus constitute a primary source that may be damaged or destroyed, but is not prone to textual alterations. However, this only applies to the object as such. A comparatively large portion of the Ancient South Arabian material is only known from copies or transcriptions of various qualities, made by modern scholars. As photographic documentation, though already requested in the review literature at a comparatively early time (cf. Schlobies, 1936, p. 58, n.1), is often lacking or of poor quality, the actual appearance of these inscriptions can no longer be checked. This material contains a certain amount of corrupt forms, including obvious faults (cf. Stein, 2002, esp. the exhaustive examples pp. 447–452). The latter are often corrected by later reeditions based on photographs or rediscovered originals.⁴ For a large amount of poorly published material there is still no reliable documentation available. Nevertheless, rectification of obvious faults was often undertaken by later editions. As a result, these inscriptions present a certain amount of variant readings that can rather be compared to manuscripts.

² These function in a roughly similar way to spacing in modern Arabic, i.e. there is a certain range of prefixes, mainly monoliteral particles, and suffixes, mainly pronominal suffixes, that are supposed to form a continuous string of characters with the main word.

³ This only applies to the very end of a string of characters (see above), long vowels before suffixes always being suppressed in script.

⁴ A particularly prominent example is Ir 13 published by al-'Iryānī (1973, pp. 74–86) on the basis of a highly defective copy. A portion of the text has been rediscovered and is reedited by Arbach (2001). Certain grammatical and lexical oddities disappeared in the course of this reedition, e.g. an ungrammatical infinitive *qtq* in § 5 that turned out to be a simple *qtl* (cf. the reedition, line 4). Note that the misread form is included in both Beeston et al. (1982, p. 110) and Biella (1982, p. 455), and has thus become part of lexicographical discussion.

9.2.2 Collection of Material

Ancient South Arabian inscriptions are scattered over a wide range of different publications which may comprise huge collections of different material (e.g., CIH and RES) and full inventories of excavations (e.g., Jamme, 1962), but may also focus on individual texts or passages. An exhaustive collection of material is undertaken by DASI, but is not yet completed for Sabaic. The compilation of material underlying the present dictionary already started back in the late 1990s, then still in a DOS format. This collection consisted mainly of analytic transcriptions and bibliographic information. The material was originally meant as a basis for grammatical studies and was thoroughly prepared for this purpose. Information on Semitic roots or fuller, non-assimilated forms was thus encoded. Since the latter are also important for lexicographic work, the compiled file, subsequently completed and augmented by newly published material, was considered an ideal base for a dictionary. The data is by now converted in an XML format.

9.2.3 Organisation of Material

The dictionary has a modular approach. The Sabaic material is split into several sub-corpora (such as votive inscriptions, building inscriptions, juridical texts etc.), which are processed separately. For the time being, the dictionary includes major parts of the votive inscriptions, which actually form the most comprehensive genre among the Sabaic texts (an up-to-date account of the incorporated material can be found on the website).

Within these corpora, lexicographic work is organized according to inscriptions, i.e. all lexical items of a given text are considered, irrespective of alphabetical order. Work thus started with a micro-dictionary containing the vocabulary of a single inscription, this core being constantly enlarged with material from other texts. Therefore, not only does the actual number of lexemes increase over time, but also their extent.

The project aims to present all extant lexical material. This also includes probable or even actual faults. Variant readings and – to a lesser extent – unusual or even faulty orthographies are thus included in the basic text.⁵ However, uncertain and incorrect readings are marked as such. A reconstruction of “correct” texts is intended for presentation purposes, but often turns out to be impossible. This is, in most cases,

⁵ Note that the textual structure of passages with variants is fairly complicated in respect to encoding. This unfortunately leads to unexpected and unwanted results that also effect the presentation in the website.

due to deficient editions lacking proper documentation (cf. section 9.2.1), but may also result from the limited range of our knowledge, especially in damaged contexts.

9.3 Morphological Analysis

The morphological analysis of texts is performed by manipulating the XML document. In the process of tagging a word, information is stored as XML attributes to the corresponding XML tag. These attributes provide information on the actual lemma (e.g. *bn* “son” vs. *bn* “bān-tree”, preposition *bn* “from” etc.), its lexical category and grammatical subcategories (if appropriate). Lexical categories include noun, pronoun, verb, preposition, conjunction, other particle and other.⁶ To facilitate a clear presentation of the lexical material, names⁷ and fragments⁸ are treated as separate categories. Since this classification is part of the lemma, as are roots and meanings, it is substituted automatically once the correct lemma is chosen. In ambiguous contexts, multiple tags can be assigned. However, this possibility is kept to a minimum to avoid confusion of the reader.

On the other hand, grammatical subcategories such as gender, number, state, conjugation and the like, are specific to the actual word. Since the defective Sabaic orthography includes many ambiguous forms⁹, these tags are edited manually, based on the particular context of the word. If neither form nor context allows a clear identification of categories, forms are tagged as “unspecific”.¹⁰ Morphological tags are used to create a morphological catalogue of attested forms, which is a vital part of the reference dictionary.

⁶ This category – in reality almost empty – was created to avoid technical problems with unexpected cases.

⁷ These show substantial differences from “normal” lexemes. While forms like *yzd* or *yhn'm* expose a difference between (verbal) morphological form and (nominal) syntactical content, composite names like *krb-'l* or *whb-'wm* cannot be assigned to a single root. A full inclusion of names would thus destroy the otherwise exclusive relation between lemma and root. Furthermore, a thorough analysis of onomastics is not in the scope of the present project. Nevertheless, different categories of names (such as anthroponyms, theonyms, toponyms etc.) are differentiated so as to facilitate further studies on this material.

⁸ This only applies to forms that cannot be reconstructed with a sufficient degree of certainty. A comprehensive study of the latter is not appropriate at the given state of research. Furthermore, minor fragments, often only consisting of a single initial or final letter preserved, are highly ambiguous and do not constitute lexemes in the proper sense of the word.

⁹ Thus, a form like *hmr-hw* can be interpreted as an infinitive “to grant him” and as a finite form “he granted him” or “they granted him”.

¹⁰ A general use of multiple tags in these often fragmentary contexts would lead to a rather unwelcome artificial increase of the catalogue of forms and thus does not allow adequate presentation.

Furthermore, tagging provides information on reliability of attestation: certain, uncertain, supplemented¹¹ or wrong¹². While both uncertain and wrong forms are marked as such,¹³ supplemented forms are simply excluded from presentation.

9.4 Definition of Lemmata

The definition of lemmata follows practical considerations. Forms are thus treated as separate items if their respective contexts show sufficient differences to consider them as such. Sabaic being an extinct language, this is more or less the approach of all extant lexicographic literature in the field.

9.4.1 Treatment of Homographs

Homographs belonging to different grammatical categories are commonly treated as separate items in the scientific literature (cf. the organization of material in both Beeston et al., 1982 and Biella, 1982). Obviously, a noun like *qm* “garrison” should be differentiated from the homographic verb *qm* “to be on garrison duty”. The difference between such morphological categories is, however, sometimes difficult to observe in existing contexts. Distinguishing between infinitives of the base stem, following a pattern *fʿl*, and abstract nouns showing a similar pattern in the construct state, is a particularly delicate case.¹⁴ Particularly in stereotype contexts, the grammatical form of a certain lexeme may resist disambiguation.

The situation is even more complicated for homographs belonging to the same grammatical category, i.e. homograph verbs or homograph nouns. These were probably differentiated via vocal patterns, a means that is not featured in the defective script (for nominal forms cf. Stein, 2003, esp. pp. 56–62). Furthermore, the rather complicated system of verbal stems and their eventual graphic distinction was only fully understood in the last decade (cf. Multhoff, 2011). Consequently, Sabaic lexicography has not yet developed consistent standards to deal with this material. Delimitation of lexemes in existing dictionaries is thus often rather arbitrary,

¹¹ Forms are generally classified as “supplemented” if no single radical is preserved.

¹² This includes both misreadings by modern editors and actual mistakes by the Sabaean writer. A further specification, though desirable as such, is in many cases impeded by a lack of reliable (photographic) documentation.

¹³ Uncertain readings are marked with a minor question mark (usually only if the reading is certain neither from script nor from context); incorrect forms and identifications are crossed out.

¹⁴ An example is *šy* “upheaval” in the expression *nqʿ w-šy šnʿm* “elevation and upheaval of an enemy”: the word could well be defined as an infinitive of the equally attested verb *šy* of more or less equivalent meaning.

sometimes even summarizing different graphemes under one single lemma.¹⁵ Given the progress in our understanding of Sabaic verbal stems over the last fifteen years, it is now possible to distinguish different verbal stems. All stems show at least occasionally unequivocal forms, mostly in infinitives. Even though these are not always attested, the fundamental disambiguation of the system has yielded a set of semantic criteria that often enables definition in otherwise uncertain cases. Verbal homographs can thus normally be clarified, presenting different stems as different lexemes and identical stems as singular lexemes, as with *ʿrb* *O*₁ “to enter” besides *O*₂ “to offer” (but see below, section 9.4.2).¹⁶

The situation is more complicated with nominal forms. If we compare related languages such as Arabic, a much bigger amount of homographs is to be expected, but clear morphological or semantical criteria are missing. In this particular case, contexts are carefully checked. Forms appearing in similar contexts are generally treated as a single lexeme. Other forms are considered as such if there are no convincing arguments against this assumption. On the other hand, forms are split into different lexemes if clear semantic differences appear from constructions or contexts, as in the case of *ḏhb*, a homograph comprising lexemes with the meaning “bronze”, “oasis, irrigated area”, “irrigation, flood-water” and a certain “measure of capacity”.

9.4.2 Deliberate Splitting of Lexemes

In certain cases, single lexemes are deliberately split up. The most common cases are particles. Conjunctions are generally separated from homograph prepositions. Very frequent particles such as *b-*, *l-* and the pronoun *ḏ-* are further divided into different semantic or contextual categories to enable a clear presentation.¹⁷ Verbal and nominal lexemes are split if they are clearly derived from different forms.¹⁸ Lastly, a small

¹⁵ Cf. e.g. the merge of *hḏrʿ*, *tḏrʿ* and *stḏrʿ* into a single entry that can be observed in Beeston et al., 1982, p. 42. A critical re-evaluation of this approach can already be found in the review by Lundin (1987, p. 49): “As a result, the *lemmata* [sic. l.] are presented in far too abstract a way without distinguishing words with similar meanings and grammatical forms (i.e. verbal stems).”

¹⁶ However, uncertain cases such as morphologically ambiguous forms in divergent contexts often have to be considered as separate items.

¹⁷ This is often done in dictionaries containing selected examples (as is the case with Biella, 1982), but avoided in more concise presentations such as Beeston et al., 1982. However, this approach proved rather dysfunctional in practice for most lexemes concerned in the context of a full citation of references.

¹⁸ Thus *hwfy* is split into denominal “to give well-being” (from *wfy* “well-being”) and deverbal “fulfil, hand over” (from *wfy* “to belong to”).

number of nouns in ubiquitous contexts are given separate entries to allow a clearer presentation of the remaining forms in the context of a fully referenced dictionary.¹⁹

9.4.3 Heterographs with Identical Meaning

Sabaic shows a certain number of words that are different in form, but apparently similar in meaning. This mainly applies to a) different rendering of weak radicals (w or y) and b) otherwise identical nouns with and without final *-t*. While some of these forms can be explained by diachronic or regional variation,²⁰ the motivation of other variant forms is less clear.²¹ Those may represent different lexemes, but may also refer to different numbers (the distribution of which rests equally unclear). Different graphemes are considered as one single lemma if their relation is clearly grammaticalized, as is the case with diachronic or regional variation. Other cases are mostly treated as different lexemes.²²

9.4.4 Treatment of Incorrect Forms

The published Sabaic material, accumulated over a period of almost 150 years, comprises a surprisingly high number of incorrect forms. These include both actual faults of the Sabaean writer (e.g. a merger of similar characters, as in *'rz* instead of *'rḏ* “earth”) and misreadings (sometimes even misspellings) by modern editors. In the absence of reliable photographic documentation, it is difficult, if not impossible, to attest to the correctness of a given form.

However, misreadings have sometimes become real “classics” over a certain period of time and could have provoked a rather huge amount of material, both translations and further reflections, and even found their way into the extant dictionaries.²³ A simple exclusion of these ghost-words from the dictionary will probably not solve the problem of their constant reappearance in (especially non-specialist) literature. On

19 This is the case with *b'l* “lord” in the phrase “NN [theonym], the lord of NN [toponym]” and *mlk* “king” in the phrase “NN, the king of Saba' [and ...]”. In the latter phrase, all ethnonyms concerned, (i.e. Saba', *ḏū-Raydān*, *Ḥaḏramawt* and *Yamanat*) are also split up.

20 Cf. e.g. verbal roots III w that are generally rendered as III y in Southern Sabaic dialects and late Sabaic texts (the latter actually being an offshoot of South Sabaic), thus *ḏky* besides *ḏkw* “to send; to expel”.

21 Cf. e.g. *b's* besides *b'st*, both meaning “evil, malice”.

22 In some uncertain cases, however, multiple tags may be used to link several possible lexemes to one single form.

23 Cf. e.g. **mhrk* “booty” featured in Beeston et al., 1982, p. 57, and Biella, 1982, p. 117, misread from damaged *mhr̥g*.

the other hand, assumed scribal faults are not always as obvious as the example given above. At least part of this material may, in fact, prove correct if further documentation becomes available. And then there are textual emendations in older editions, the status of which is often rather unclear. All these forms are therefore to be included – and properly commented – in the dictionary. And finally – lemmata, which are based on actual misread forms and are thus to be deleted from the present corpus, may in fact appear as clearly attested forms in the future.

9.5 Presentation of Material

9.5.1 Structure of Presentation

While the actual processing work is structured by inscriptions (see section 9.2.3, above), the presentation of the lexical material in the dictionary is generally structured by lexemes. Since Sabaic grammar does not really match the internal logic of alphabetical arrangement, ensuring usability proved rather tricky. While lexemes are operated in a standardized form for both internal and presentational purposes, this form is often difficult to reconstruct from an internal plural form²⁴, or an irregular formed verbal stem²⁵, and is thus not a reliable basis for arrangement or even search. Dictionaries for other languages with similar phenomena (such as Arabic or Gə‘əz) often opt for an arrangement by root. The latter, however, may also prove difficult to reconstruct and is thus an equally unreliable basis for search. We therefore decided to offer several different search options to ensure easy access to the material: lexemes proper, but also roots, strings of characters and translations.²⁶ All lexemes are complemented with a suggested translation, an automatically generated counter giving the number of attestations in the material processed thus far, a catalogue of attested morphological forms, a complete literary and etymological documentation and quotations of the particular lexeme in its syntactical and semantic context (Figure 9.1). For the time being, the presentation is exclusively in German. However, technical requirements for an eventual extension to other languages were taken into consideration.

²⁴ Such as *’byt* “houses”, sg. *byt*, or *mšymt* “(agricultural) installations”, sg. *mšm*.

²⁵ Such as *htrgn* (infinitive) besides *thrg* (suffix conjugation) “to fight”.

²⁶ I.e., the German translations suggested in the dictionary.

Suchergebnis für sb' 01 131 Belege in 86 Texten

Übersetzung [Schrift verkleinern](#)

ausziehen, aufbrechen, zu Felde ziehen

Ältere Übersetzung [Aufklappen](#) [Schrift verkleinern](#)

(bellum) gerere
CIH I, 408
(eine Reise) unternehmen
Sims 2000-24 Rev. 8

Altsüdarabische
Parallelen [Aufklappen](#) [Schrift verkleinern](#)

Qatabanisch
sb'
(einen Feldzug) unternehmen
Lemus 1098-272

Etymologische
Parallelen [Aufklappen](#) [Schrift verkleinern](#)

Arabisch
masba' (Wz. sb') "wa-l-masba'u: t-tariqu fi l-ğabali 'masba' ist der Weg im Gebirge" Lisān III / 227
sub'a (Wz. sb') "wa-'innaka la-turidu sub'atan 'ay turidu safaran ba'idan yuğayyiruka 'du

Formen [Aufklappen](#) [Schrift verkleinern](#)

SK
3. m. sg. s[b']
CIH 312/4?
3. m. sg. sb'

Wendungen [Zu klappen](#) [Schrift verkleinern](#)

sb' 'dy "ausziehen nach"
b-kn sb' 'dy shrtm, Ry 538/14.-15., "als er nach Sahiratān zog"
b-qt hwfy-hmy 'lmqh b-kn sb'y 'dy kbtm b-wrh q-'lt q-hrf 'b'mr bn wdd'l bn ḥzfm q-ḍmrn,
Schm/Sir 133/4.-8., "dafür, daß 'Almaqah ihnen Wohlergehen geschenkt hat, als sie im

Figure 9.1: Result for lemma. The example is sb'

9.5.2 Accessible Material

9.5.2.1 Translation

An up-to-date translation is given. This is established on the basis of the complete epigraphic material. To enable retrograde search and thus usability, it is often enriched with synonyms. In ambiguous cases, all possible renderings are mentioned, including possible references to other lexemes.²⁷ Onomastic material is generally vocalized to facilitate reading. It should be kept in mind, however, that most of this vocalization is conventional.

²⁷ This is especially important for supposedly incorrect forms.

9.5.2.2 Existing Translations

A full catalogue of existing translations is being prepared.²⁸ This is particularly important since the meaning of many lexemes is still not sufficiently established in the literature. The whole range of possible interpretations should thus be made accessible to the user to reflect scientific discussion in the field. Part of the collected material stems from existing dictionaries such as Beeston et al., 1982, and glossaries to larger corpora such as Jamme, 1962, pp. 426–451. Unfortunately, the latter are not always comprehensive.²⁹ However, these sources in no way reflect the totality of existing interpretations. Further material is retrieved from editions (usually containing translations in context) and commentaries. In particular, translations referred to in commentaries to text editions are often dispersed and thus difficult to access.³⁰ Furthermore, existing translations of texts are thoroughly checked for their lexicographic content.³¹ However, to allow a satisfactory workflow, translations have to be near to literal. Paraphrases, especially common for phrases considered as hendiadys, tend to be tricky. Nevertheless, these are included if they can be linked with certainty to particular lexemes.

The resulting catalogue is often surprisingly extensive. Since sufficient criteria for classifying variant forms as “identical” are missing, only verbatim quotes are considered as a single item. Catalogue entries may thus be rather close to each other. In exceptional cases such as *sb'* “to go out”, approximately one hundred different translations have been collected, reaching from “(bellum) gerere” up to “zum Kriegszug aufbrechen”. While translations of common words normally started to center around a semantic nucleus at a comparatively early stage, sometimes back in the 19th century, translations of other lexemes can differ considerably over time. In particular, translations of forms with rather unspecific literal meanings, such as *sb'*, may also include a wide range of metaphorical renderings oriented towards context rather than literal meaning.

In the case of onomastics, all existing vocalizations are collected. As in the case of other lexemes, only verbatim quotes were subsumed under a single entry. However, this approach proved rather dysfunctional, given the huge amount of different

28 Though some dictionaries (such as Biella, 1982 and Ricks, 1989) are commonly thought to be rather compilations of extant material than independent lexicographical work, this has never been done explicitly for the whole lexicon.

29 This is e.g. the case in Beeston, 1976 (pp. 60–72), where only a “select glossary” is presented that is incomplete both in terms of lexemes and translations.

30 Thus Ryckmans (1964, p. 91) has stated in his review of Jamme, 1962 “Ces suggestions ne manquent pas toujours d’intérêt en soi, mais outre qu’elles sont disséminées dans tout l’ouvrage, et par conséquent pratiquement perdues [...]”. It should be noted that commentaries often refer to lexemes not included in the commented text.

31 This is mostly done while preparing the inscription that is currently processed. Translations from inscriptions which have not yet been processed are only sporadically included.

transliteration systems that entails numerous pseudovariants (e.g. Dât-Ḥimyam besides dhāt Ḥimyam, the name of a female deity).

9.5.2.3 Etymological Parallels

Lemmata are enriched with etymological parallels from South Arabia and beyond. These are divided in non-Sabaic Ancient South Arabian, i.e. Qatabanic, Minaic and Ḥaḍramitic, and other Semitic languages. Ancient South Arabian parallels are catalogued according to Sabaic, i.e. a catalogue as complete as possible is intended. This includes material from both dictionaries and translations of actual texts, irrespective of correctness.

Since scientific literature (at least up to the first half of the 20th century) normally considered the different idioms as mere variants of a common Ancient South Arabian language, translations were often meant as applying to all respective languages. They can thus be considered as part of the catalogue of existing translations for Sabaic as well.³² Etymological parallels from other Semitic languages are generally retrieved from the respective dictionaries. Cataloguing started with geographically adjacent languages (Arabic, Ethiopic and Modern South Arabian) and is by now far from complete, but is continuously being enriched.

9.5.2.4 Morphological Catalogue

An exhaustive morphological catalogue is created from the morphological tags. The section gives a fully referenced overview of attested forms,³³ which is arranged according to morphology.³⁴ All references are clickable and linked to quotations in context.

9.5.2.5 Examples in Context

The actual usage of each particular lexeme is illustrated by references within their contexts.³⁵ These are given both in transliteration of the Sabaic text and full German

³² In some cases, translations in the secondary literature cannot be classified by language. These were normally taken as Sabaic.

³³ A complete catalogue of evidence is given in no other dictionary on the subject. This is, however, probably rather due to limited space and high printing costs than to editorial choices.

³⁴ Due to technical algorithms of the programme, incomplete forms from damaged passages split by brackets in the transliteration have to be given as separate entries. The catalogue of suffix conjugation forms of the 3rd person masculine singular of *hmr* “to grant” thus features by now a total of 8 different strings, including also entries like *hmr*], *hmr*, *hmr*], *h*[*mr*, and so on besides complete *hmr*.

³⁵ Examples in contexts are given in Biella, 1982, though the collection is far from being exhaustive. The more concise presentation in Beeston et al., 1982 is devoid of such material.

translation. In order to demonstrate the different semantic and/or syntactical aspects of a word in different contexts, these quotations are structured by significant headlines that provide an overview of usages. For extensively attested lexemes,³⁶ however, even this classification will fail to guarantee a fairly clear presentation.

To ensure a consistent rendering and improve workflow, quoted passages are only translated once. Inscriptions are therefore split into paragraphs covering sufficiently comprehensible semantic units³⁷ and supplemented with a German translation. Renderings are kept as literal as possible. As all elements of texts are stored separately in the database, each chosen paragraph can be created using simple routines. The paragraphs thus created are subsequently allocated to their correct place in the structure of presentation of each single lexeme.

9.6 Results Reached Thus Far

The web presentation of the project, accessible under [<http://sabaweb.uni-jena.de>], was launched in 2016. At present (June 2018) it contains over 1,800 lemmata³⁸ (plus over 2,200 names) attested in around 900 inscriptions containing a total of 70,000 words. Choosing digital technology in preparing and presenting the dictionary certainly had substantial effects on the workflow. Huge amounts of diverging material such as inscriptions on the one hand and translations and etymologies on the other are easy to manage in the framework of a database. Lexicographical work can be structured according to internal criteria such as contexts and does not depend on external necessities such as alphabetical arrangement. The same applies to presentation, as lexemes can be published in any possible order without affecting usability.

Acknowledgements: The work presented in this present paper is, of course, not entirely my own. The team is made up of Norbert Nebes (project director), Anne Multhoff (responsible editor), Mariam Kilargiani (editor) and Heiko Werwick (software engineer). External cooperation partners are Ingo Kottsieper, Göttinger Akademie der Wissenschaften, and Peter Stein, Friedrich-Schiller-Universität Jena – Theologische Fakultät (minuscule inscriptions). The project started in 2012 and is scheduled for nine years. It is hosted at the Friedrich-Schiller-Universität Jena and generously funded by the Deutsche Forschungsgemeinschaft.

³⁶ Such as *hqny* “to dedicate”, showing more than 500 attestations. Since the project aims at a complete documentation of all attested material, it is almost impossible to keep the structure for such huge amounts of material visible.

³⁷ Though these may be complete sentences, the specific syntactical structure of Sabaic inscriptions often requires paragraphs on a sub-sentence level.

³⁸ This refers only to material publicly accessible. By now, basic information on over 4,000 different lexemes has been collected.

Bibliography

- Arbach, M. (1993). *Le maḏābīen: Lexique - Onomastique et Grammaire d'une langue de l'Arabie méridionale préislamique. Tome I. Lexique maḏābīen. Comparé aux lexiques sabéen, qatabānite et ḥaḍramawtique* (PhD thesis). Aix-en-Provence: Université de Provence Aix-Marseille I.
- Arbach, M. (2001). Une photographie inédite de l'inscription Ir 13. *Raydān*, 7, 13–24.
- Arbach, M. & Schiettecatte, J. (2006). *Catalogue des pièces archéologiques & épigraphiques du Jawf au musée national de Ṣan'ā'.* Ṣan'ā': Centre français d'archéologie et de sciences sociales de Ṣan'ā'.
- Beeston, A.F.L. (1976). *Warfare in Ancient South Arabia (2nd. - 3rd. centuries AD)*. London: Luzac.
- Beeston, A.F.L., Ghul, M.A., Müller, W.W., & Ryckmans, J. (1982). *Sabaic Dictionary (English-French-Arabic)*. Louvain-la-Neuve: Peeters / Beyrouth: Librairie du Liban.
- Biella, J.C. (1982). *Dictionary of Old South Arabic. Sabaean Dialect*. Chico, CA: Scholars Press.
- CIH. *Corpus Inscriptionum Semiticarum. Pars quarta: inscriptiones ḥimyariticas et sabæas continens*. Paris: E Reipublicæ typographeo.
- al-'Iryānī, M.A. (1973). *Fī tāriḥ al-Yaman. Ṣarḥ wa-ta'liq 'alā nuqūṣ lam tunṣar. 34 naqṣan min maḡmū'at al-Qāḏī 'Alī 'Abdallāh al-Kuhālī. al-Qāhira*.
- Jamme, A. (1962). *Sabean Inscriptions from Maḥram Bilqīs (Mārib)*. Baltimore, MD: John Hopkins Press.
- Jamme, A. (1985). *Miscellanées d'ancien arabe XIV*. Washington [privately printed].
- Lundin, A.G. (1987). Sabaean Dictionary. Some Lexical Notes. In C. Robin & M. Bâfaḡīh (Eds.), *Ṣayḥadica. Recherches sur les Inscriptions de l'Arabie Préislamique Offertes par ses Collègues au Professeur A.F.L. Beeston* (pp. 49–56). Paris: Geuthner.
- Maraqten, M. (2014). *Altsüdarabische Texte auf Holzstäbchen. Epigraphische und kulturhistorische Untersuchungen*. Würzburg: Ergon.
- Multhoff, A. (2011). *Die Verbalstambildung im Sabäischen* (PhD thesis). Jena: Friedrich-Schiller-Universität Jena.
- Priolella, A. (2013). *Inscriptions from the southern highlands of Yemen. The epigraphic collections of the museums of Baynūn and Dhamār* (Arabia Antica 8). Roma: L'«Erma» di Bretschneider.
- RES. *Répertoire d'Épigraphie Sémitique. Publié par la commission du Corpus Inscriptionum Semiticarum*. Paris: Imprimerie nationale.
- Ricks, S.D. (1989). *Lexicon of Inscriptional Qatabanian*. Roma: Editrice Pontificio Istituto Biblico.
- Robin, Ch.J. (2013). À propos de Ymnt et Ymn: “nord” et “sud”, “droite” et “gauche”, dans les inscriptions de l'Arabie antique. In Fr. Biquel-Chatonnet, C. Fauveaud, & I. Gajda (Eds.), *Entre carthage et l'Arabie heureuse. Mélanges offerts à François Bron* (pp. 119–140). Paris: De Boccard.
- Ryckmans, J. (1964). Review of A. Jamme, *Sabaeen Inscriptions from Maḥram Bilqīs (Mārib)*, with foreword by Wendell Phillips. Baltimore, The Johns Hopkins Press, 1962 (...). *Bibliotheca Orientalis*, 21, 90–94.
- Schlobies, H. (1936). Neue Dokumente zur altsüdarabischen Epigraphik. *Orientalia*, 5, 57–63.
- Sima, A. (2000). *Tiere, Pflanzen, Steine und Metalle in den altsüdarabischen Inschriften. Eine lexikalische und realienkundliche Untersuchung*. Wiesbaden: Harrassowitz.
- Stein, P. (2002). Schreibfehler im Sabäischen am Beispiel der mittelsabäischen Widmungsschriften. *Le Muséon*, 115, 423–467.
- Stein, P. (2003). *Untersuchungen zur Phonologie und Morphologie des Sabäischen*. Rahden/Westf.: Marie Leidorf.

- Stein, P. (2010). *Die altsüdarabischen Minuskelinschriften auf Holzstäbchen aus der Bayerischen Staatsbibliothek in München. Bd. 1. Die Inschriften der mittel- und spätsabäischen Periode.* Tübingen/Berlin: Wasmuth.
- Stein, P. (2011). Ancient South Arabian. In S. Weninger (Ed., in collaboration with G. Khan, M.P. Streck, & J.C.E. Watson), *The Semitic Languages. An International Handbook* (pp. 1042–1073). Berlin/Boston: De Gruyter.