

## THE SWAHILI MANUSCRIPTS PROJECT AT SOAS

YAHYA ALI OMAR AND ANNMARIE DRURY<sup>1</sup>

Researchers in the field of Swahili studies have noted the need for a comprehensive, detailed catalogue of the Swahili manuscripts at the School of Oriental and African Studies (SOAS, University of London). They have had occasion to make this recommendation in print (see, for example, Omar and Frankl 1994 and Frankl 1999) and – presumably many more times, and to no specified audience – in life, as they pored over unitemised or inaccurately labelled files of eminent scholars in the field. It is precisely this need for a comprehensive, widely accessible catalogue that the Swahili Manuscripts Project at SOAS was established, in October 2001, to address.<sup>2</sup> In this article, the present authors undertake to describe the catalogue being created and to discuss problems they have encountered and observations they have had opportunity to make in the course of their work.

### Nature of the collections and catalogue

The SOAS archives hold seven major Swahili collections: that of the Rev. W.E. Taylor (the Taylor Papers), of Alice Werner (the Werner Collection), of William Hichens (the Hichens Collection), of J.W.T. Allen (the Allen Collection), of Wilfred Whiteley (the Whiteley Collection), of Jan Knappert (the Knappert Collection) and, begun in autumn 2001, of Sheikh Yahya Ali Omar (the Y.A. Omar Collection). The manuscripts date from the 1790s to the 1970s, with the approximate dates of manuscripts in the Taylor Papers being 1810-1899; the Werner Collection 1906-1934; the Hichens Collection 1792-1943; the Allen Collection 1898-1977; the Whiteley Collection 1950-1969; the Knappert Collection 1929-1972; and the Y.A. Omar Collection 1960-1973. These ranges represent the dates of the creation of manuscripts and in many instances, of course, differ from the dates when the contents of manuscripts were composed. The two manuscripts of the *Hamziyya* in the Hichens Collection, for instance, dated 1936 and 1792, both contain the Swahili poem composed in 1749.

---

<sup>1</sup> The authors of this article must recognise the involvement of many in the project for which they are the two designated 'researchers': Dr Ridder Samsom, now of Humboldt University, who conceived the project; Professor Graham Furniss and Dr Farouk Topan of the SOAS Africa Department, and Mrs Rosemary Seton of the SOAS archives, who comprise the 'project management group'; and Ms Rhiannon Stephens, now a PhD candidate in the history department at Northwestern University, who completed much early work on the catalogue. Technical support is provided by Mr Junaid Minar of the IT Department at SOAS.

<sup>2</sup> The Swahili Manuscripts Project is funded by a 31-month research grant from the Leverhulme Trust. Prior to the awarding of this grant, the project received from SOAS funding for a start-up period during which software was tested and a catalogue prototype was designed.

To make details of the Swahili manuscripts, the vast majority of which originated in regions that now comprise Kenya and Tanzania, available to a large international audience of scholars, the project team is publishing the catalogue on the internet. Accompanying the web version of the catalogue will be manuscript images and recordings of Yahya Ali Omar reciting selected materials. The project will also produce a printed volume. In creating the catalogue, the researchers' primary aim is to develop content descriptions that will assist users firstly in understanding abstruse or challenging material, particularly poems; and secondly in distinguishing, even from afar, the various extant versions of a single poem, such as the *Utenzi wa Barasisi* (at least five versions in the SOAS collections alone) or the *Utenzi wa Mikdadi* (at least two versions). In this sense, the researchers intend to move a stage beyond a catalogue such as J.W.T. Allen's of the manuscripts at the University of Dar es Salaam (1970) by presenting substantial content descriptions while maintaining scrupulous attention to other details. The physical traits of manuscripts, such as size and type of paper, ink colour and the use of illumination in scribing, are not emphasised to the degree that they are, for example, in Adam Gacek's *Catalogue of the Arabic Manuscripts in the Library of the School of Oriental and African Studies, University of London* (1981). In all instances the attempt is made to privilege content, with content descriptions based upon the expertise of the lead researcher, Yahya Ali Omar.

The functional basis of the catalogue is of course a database, in this instance a hierarchical database comprised of three levels: 'collection,' 'file' and 'item.' Eight entries at the collection level correspond to the seven named collections and a miscellaneous collection. At the file level, entries correspond to the numbered and titled files listed in the archives accessions register. Many files, while physically comprising a single folder or notebook, contain a number of separate and distinct writings, and the item level was established to treat these writings individually. Thus, for example, manuscript 380066, entered in the accessions register as the *Utenzi wa Muhammad Kijumwa Kumuusia Mwanawe Helewa*, in fact contains three complete poems (the *utenzi* and two *mashairi* from Lamu) and one poetic fragment (opening stanzas of the *Utenzi wa Ndiwa na Kozi*). In the many instances like this, each poem within a file is assigned its own item number (380066a, 380066b and so forth) and catalogued as a separate entity; the title of the file is also changed to better reflect the contents.

At each of its three levels, the catalogue is comprised of 'fields' that represent the subjects for which data is presented. The researchers used the standards of the International Council on Archives (ISAD (G), second edition) as a basis for the selection of fields.<sup>3</sup> They made alterations and additions, however, to reflect exigencies of the material and the anticipated interests of users. The result is a slightly specialised catalogue, with specialisation being evident at the most specific level of 'item.' Thus while the fields at the collection level

---

<sup>3</sup> The standards (1999) may be accessed from the internet at <http://www.ica.org/eng/mb/com/cds/descriptivestandards.html>. See bibliography for details of the print version (2000).

(comprising 'title of collection,' 'collection reference,' 'dates,' 'relevant publications,' 'biographical information' and 'notes') and at the file level (including 'title of file,' 'file reference,' 'content,' 'source of acquisition by SOAS' and 'notes') are quite usual, the item level contains 'relevant dialects' and 'scribe' fields that deviate from general archival usage. Both in number and nature, the fields at item level reflect the aim of describing aspects of the manuscripts in which users are expected to have the greatest interest. In addition to the fields identifying an item by title and by the file and collection to which it belongs, and to the 'scribe' and 'relevant dialects' fields mentioned above, the item level contains the following: 'first lines of manuscript'; 'author(s)'; 'scribe'; 'AD date'; 'AD date of composition'; 'AH date'; 'AH date of composition'; 'extent,' meaning length or size; 'resource type,' of which more below; 'poetic form,' also discussed below; 'format,' whether handwritten, typed or a combination of the two; 'language,' whether Swahili, English or another; 'script,' whether Roman or Arabic; 'subject and keywords,' with keywords taken from the *UNESCO Thesaurus*;<sup>4</sup> 'people' and 'places,' to allow search by proper nouns relevant to Swahili studies; 'biographical history'; 'archival history'; 'physical characteristics'; 'existence/location of copies'; 'relevant publications'; 'notes'; and 'scope and content.' References to facsimiles and digital recordings are also included.

In planning the catalogue at item level, the authors devoted special attention to the 'resource type' and 'poetic form' fields, because these relate to the classification of literary genres in Swahili. For the sake of indexing and consistency, a limited list of possible inputs was designed for these fields. Creating that list involved two objectives: to be specific while also allowing for variation among the manuscripts; and to conform both to the nature of the manuscripts and to the natural expectations of researchers in the field. Because scholars have proposed a variety of systems for classifying poetic form in Swahili, the authors decided that the catalogue would profit from consultation in this area.

Thus the manuscripts project invited Abdilatif Abdalla of the University of Leipzig and S.A.K. Mlacha of the Institute for Kiswahili Research at the University of Dar es Salaam to SOAS to discuss questions of genre and poetic form. During two days of consultation, participants referred to writings on prosody by Ibrahim Noor Shariff (1988), Al-Amin bin Ali Al-Mazrui (nd), K. Amri Abedi (1954) and William Hichens and Mbarak Ali Hinawy (1963). They also examined examples of manuscripts from the SOAS collection and discussed the structure of the database. In conclusion, they agreed that the categories for 'resource type' should be 'letter,' 'poem,' 'essay,' 'notes,' 'story,' 'song,' 'maulid,' 'Qur'anic translation/exegesis' and 'various'; and that the categories for 'poetic form' (applicable only if the 'resource type' is 'poem') should be 'utenzi,' 'utumbuizo,' 'shairi,' 'shairi: takhmis,' 'shairi: wimbo' and 'various.' The classifications of maulid and Qur'anic translation/exegesis

---

<sup>4</sup> The authors have used the web version available at <http://www.ulcc.ac.uk/unesco/>. A reference for the print version (1995) appears in the bibliography.

reflect the existence of a significant number of such manuscripts in the SOAS collections. Some complications surrounded the use of 'song' and 'wimbo' categories. In general, it was decided that 'song' in 'resource type' should be used for materials other than three-line *nyimbo* – such as the songs of children at play, songs associated with marriage festivities, or a lullaby. The classification 'shairi: wimbo' should be used for materials that exhibit the 'wimbo' form as described by Shariff (1988) or Mazrui (nd). Ultimately – as the above classifications suggest – it was agreed that *nyimbo* and *takhmis* should be viewed as types of *shairi*, although some participants had enduring reservations about this idea. The classifications remain flexible and may of course be modified during the process of cataloguing.

### Observations and comments

In terms of dialect, the Swahili of some manuscripts eludes simple classification, and for other manuscripts (notes by Mbarak Ali Hinawy, for instance, or an essay that won a prize from the East African Swahili Committee) such classification does not seem relevant. But with certain materials, especially with poems that consistently exhibit the characteristics of a single dialect, identification is valuable. Hence the question arises as to how to distinguish dialects from one another. For scholars who have conducted specialised study in this area, the task is relatively easy. But researchers who have not, particularly researchers who are primarily familiar with the Swahili of Zanzibar or mainland Tanzania, may find making such an identification difficult. In the course of their work, the present authors have noted a few limited but simple means of addressing such questions.

Most traditional *tenzi* and many other traditional poems will have been composed originally in a northern dialect (i.e. Kiamu, Kisiu or Kitikuu, the latter also known as Kigunya), and these dialects can generally be distinguished from southern forms through vocabulary and spelling. The use of 'y' instead of 'j,' for example, signals a northern dialect (*yana* instead of *jana*, *yuzi* instead of *juzi*), although words that have entered Swahili directly from Arabic, such as *jini*, may retain 'j.' Differences between northern and southern vocabularies are similarly easy to recognise; they include variations such as *jepa* for the southern *iba*; *taba* for the southern *igiza*; *limatia* for the southern *chelewa*; *nena* for the southern *sema*; and *mayowe* for the southern *kelele*. Another clue is provided by northern forms such as *mbee* and *nyee* for the southern *mbele* and *nywele*.

Yet while recognising a dialect as northern is fairly simple, distinguishing among the northern dialects often presents greater difficulties. One notable distinction between Kiamu and the other two, more northerly dialects lies in Kiamu's use of 'zi' – and Kisiu's and Kigunya's use of 'dhi' – where southern forms would have 'vi.' A feature that distinguishes Kitikuu from Kisiu and Kiamu is the occurrence of a dental nasal 'n' (ŋ) where other dialects

would have 'ny'; *nyama* and *nyuni*, for example, become *ṅama* and *ṅoni*.<sup>5</sup> Another distinguishing feature of Kitikuu lies in the words for 'mother' and 'father.' In Kisiu the southern *babangu*, *babako* and *babake* appear as *mbangu*, *mbako* and *mbake*; and *mamangu*, *mamako* and *mamake* similarly appear as *mmangu*, *mmako* and *mmake*. In Kitikuu, however, *babe*, *showe* and *ishe* signify *babangu*, *babako* and *babake* respectively; and *mame*, *nyawe* and *inya* signify *mamangu*, *mamako* and *mamake*. These words – although they appear in a limited number of manuscripts – do constitute, when present, a reliable indicator that Kitikuu is in use. For example, in one song among a collection of *kimai*, songs from fishing communities on the north Kenyan coast, the following lines occur:

*Babe sala na mame sala, nami mwanavao tangiya salani*  
*Babe dhicha, na mame dhicha, nami mwanavao tangiya dhichani*

Even without knowing the context of these lines, a reader can identify them as Kitikuu; in more 'standard,' southern Swahili they might be rendered:

*Babangu swala na mamangu swala, na mimi mtoto wao nitaingia katika swala*  
*Babangu vita na mamangu vita, na mimi mtoto wao nitaingia vitani*

Another song begins:

*Ho mame unambidhie kalale, kenda kalala ngomani*  
*Asubuhi kiamuka, ha mai yatele nyangwani*<sup>6</sup>

This is another example of Kitikuu; rendered in a more southern form the lines would be:

*Huyo mamangu amenambia kalale, nikenda kulala ngomani*  
*Asubuhi nilipoamka, haya maji yako tele nyangwani*

The question of dialect of course becomes complicated when material has a long history of oral transmission. Many manuscripts, particularly those of *tenzi*, in the SOAS collection exhibit northern and southern forms simultaneously. This would seem to be evidence of their movement from north to south, and perhaps of the proclivities of various scribes. One manuscript of the *Utenzi wa Abdirahamani*<sup>7</sup> for example, exhibits northern verb forms like *saa* and *poka* (for the southern *salia* and *pokonya*). 'Dh' also appears in place of 'z' in most instances: another northern characteristic. But conversely, the southern 'j' generally appears

<sup>5</sup> For further description of dialects of the north Kenya coast, see Nurse (1982). The authors thank Michael Mann for this reference and for his advice on this linguistic point.

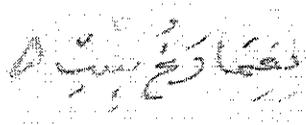
<sup>6</sup> This and the preceding extract are found in MS 380059, scribed in Arabic script by Yahya Ali Omar c. 1967. For transliterations, translations and discussions of many of the *kimai* in this manuscript, see Yahya Omar and Kevin Donnelly (1982). Italics have been added by the present authors.

<sup>7</sup> MS 255733a, scribed in Roman script by an anonymous scribe and dated 1956. Authorship of this popular traditional poem is attributed in a published edition to Hemed Abdallah Said al-Buhry, who is reported to have recollected the *utenzi* from the recitations of his grandfather. See Hemed Abdallah Said al-Buhry (1961), with J.W.T. Allen's notes on authorship.

where the northern form would be 'd' (*njia* instead of *ndia*, for example), and the scribe uses southern forms like *nchi* (rather than *nti* or *iti*) and *sogea* (instead of *songea*). Within this intriguing mix of dialectical features, we also find the *ha* verb form, an old Mombasa form: *haona*, for example, replaces *nikaona*. Such dialectical variation would seem, based on the authors' observations, to be a feature of *tenzi* manuscripts that is at least as common as, and probably more common than, dialectical consistency.

Another aspect of the manuscripts that deserves mention is the use of the Arabic alphabet. Along the southern (e.g. Kilwa and Unguja) and central (e.g. Mombasa) Swahili coast, the Arabic alphabet was the main vehicle for written Swahili prior to about 1930; on the northern coast it continues to have a predominant role. As other scholars have discussed (Allen 1945 and Angogo 1983, for example), phonological variation between Arabic and Swahili means that the alphabet of the former does not perfectly accommodate the latter. Writers address this divergence in a variety of ways. Some expand the alphabet to represent Swahili phonemes such as 'ch' or 'p,' either by devising an *ad hoc* system or by adopting one that they have encountered elsewhere: whether in manuscripts, published works or, less formally, in the writings of companions or teachers. Other scribes make do with an unmodified alphabet, leaving to the reader the task of recognising the Swahili words thus represented.

Compounding this variation in alphabet is variation in spelling. As J.W.T. Allen observes of Swahili written in the Arabic alphabet, 'there is no established spelling and each writer spells as he pronounces' (Allen 1945: 7). In most instances, a scribe's study of the Arabic alphabet will have been limited to instruction at a *chuo*, where the alphabet is taught solely to enable reading of the Qur'an, not for the purpose of learning the Arabic language *per se*, and not to enable writing in Swahili. We would also suggest that some scribes deliberately use spellings which 'sound' more Arabic, replacing the letter 'siin' with 's{aad,' for example (*s{ilaha* instead of *silaha*) and 'kaaf' with 'qaaf' (*qitabu* instead of *kitabū*). Many puzzling lines can only be comprehended through facility with the Arabic alphabet and a 'feel' for language and content. Sometimes the process is relatively simple, as with the opening lines of the *Utenzi wa Mwana Kupona*, which are well known among scholars. For example, familiarity alone would enable easy comprehension of the hemistich below; the direct transliteration is in italics.



*nighimaa waghū binti: Negema wangu binti* <sup>8</sup>

---

<sup>8</sup> This image and the two that follow appear by permission of the Library of the School of Oriental and African Studies, University of London.

With less familiar material, a reader's knowledge of alphabetical and spelling variations becomes more important, as does attention to content. The hemistich below occurs in a poem of religious admonition entitled *Waadhi wa Fuadi*.



*yiyi kiya mula waku: nyenyekea Mola wako*

Sometimes the writing of entire lines or sequences of lines must be 'decoded'. This may necessitate experimenting with various Swahili correspondences for Arabic letters, and with various meanings in a given context. A good example of an abstruse line is the one below, also from the *Waadhi wa Fuadi*.



*mtu hunina akira, asiyabaraki tara: mtu hunena akenda, asiyapanda kitanda*

In an instance like this, literary and cultural knowledge of Swahili have great value, because they assist the reader in supposing what an author would likely be saying at a given stage in a given composition.

Another factor that complicates reading is the frequent use, especially in *tenzi*, of Arabic words and phrases that have not entered Swahili as loans. Without a good knowledge of Arabic, a reader may fail to understand these words as they appear, whether in Arabic or Roman script; and even a reader with excellent knowledge of Arabic may spend several minutes confounded by the idiosyncratic spelling of a phrase. In the manuscript of the *Utenzi wa Abdirrahman* mentioned above, for example, we encounter the phrase (in Roman characters) *bitwarifati ayani*. Even with substantial knowledge of Arabic, a reader may struggle to recognise this as a Roman rendering of the Arabic *fii fwarfaḥ 'ainin*, meaning 'in the blink of an eye'. The challenges a reader faces in terms of alphabet and spelling seem to have infinite variations.

In the ten months of reading and writing about manuscripts that have preceded this article, several recommendations have evolved in the minds of the authors. Most of these, in accordance with the type of manuscript that has dominated the first portion of the project,

concern language in older Swahili poetry, especially *tenzi*. The authors would suggest, for example, that the fascinating dialectical variation evident in many manuscripts would inform study of the transmission of *tenzi*. It is precisely such an aspect of the manuscripts, an aspect that is deeply puzzling to the reader, that provides hints about the dissemination of traditional poetry and the role of individual scribes (and perhaps communities) in reworking and altering poems. The authors would also suggest that the development of a small dictionary specific to poetry that predates the twentieth century, especially *tenzi*, would be an immensely valuable tool for scholars. The language of these traditional poems, it should be remembered, presents problems not only to European and North American university scholars but also to university students in East Africa and to many of their instructors. A reference work created now, that utilises the knowledge of contemporary traditional scholars, would encourage (even enable) scholars of the future to devote attention to older traditional poetry. Two other types of reference material would be similarly valuable: one a comparative dialectical vocabulary, or dictionary, specifically referring to the contents of traditional poetry; another a work along the lines of Allen (1945) to advise scholars on the deciphering of Arabic script as it appears in Swahili manuscripts, especially the poetic.

These points, taken together, imply the authors' final recommendation. It is that much more scholarly attention be devoted to manuscripts, especially poetic manuscripts, in Swahili. Such manuscripts are often hard to read, and they complicate understanding even of canonical poems (such as the *Utenzi wa Mwana Kuponu*) that now exist in virtually canonical editions (such as Allen 1971). All of the scholars who have named Swahili collections in the SOAS archives, for example, have published materials based on those manuscripts, and their publications are well-known and often, in one way or another, viewed as authoritative. Yet in many cases, such existing publications are best thought of as representing a single, if well-informed, scholarly and editorial understanding of certain manuscript material.<sup>9</sup> Study of manuscripts tends to complicate understanding of the writings that those manuscripts contain; but this is a very necessary and good complication. It is one that we hope more researchers will seek out and will also seek, by their own well-informed expertise, to illuminate.

### Bibliography

- Abedi, Kaluta Amri. 1954. *Sheria za Kutunga Mashairi na Diwani ya Amri*, Dar es Salaam, Kampala, Nairobi: Eagle Press.
- Al-Amin bin Ali al-Mazrui. nd. "Arudhi." Unpublished essay.
- Allen, J.W.T. 1945. *Arabic Script for Students of Swahili*. Published for the editorial committee, Tanganyika Notes and Records.

---

<sup>9</sup> Poems associated with Fumo Liongo as presented by Werner (1927), Harries (1962) and Knappert (1979), all of whom made use of manuscripts now held by SOAS, illustrate this point

- Allen, J.W.I. 1970. *The Swahili and Arabic Manuscripts and Tapes in the Library of the University College, Dar-es-Salaam: a Catalogue*. Leiden: E.J. Brill, 1970.
- Allen, J.W.T. 1971. *Tendi: Six Examples of a Swahili Classical Verse Form*. London and Nairobi: Heinemann Educational.
- Angogo, Rachel M. 1983. "The Arabic Script: a Lacuna in the Study of Early Swahili Literature." *Kiswahili* 50 (2): 59-69.
- Frankl, Peter. 1999. "W.E. Taylor (1856-1927): England's Greatest Swahili Scholar." *AAP* 60: 161-174.
- Gacek, Adam. 1981. *Catalogue of the Arabic Manuscripts in the Library of the School of Oriental and African Studies, University of London*. London: School of Oriental and African Studies, University of London.
- Harries, Lyndon. 1962. *Swahili Poetry*. Oxford: Clarendon Press.
- Hemed Abdallah Said al-Buhry. 1961. *Utenzi wa Abdirrahmani na Sufiyani*. Trans. by Roland Allen, with notes by J.W.T. Allen. Nairobi: EALB.
- Hichens, William (and Mbarak Ali Hinawy). 1963. "Swahili Prosody." *Swahili* 33 (1): 107-137.
- ICA. 2000. *General International Standard Archival Description, ISAD (G), 2nd edition*. Ottawa: International Council on Archives.
- Knappert, Jan. 1979. *Four Centuries of Swahili Verse: A Literary History and Anthology*. London: Heinemann.
- Nurse, Derek. 1982. "The Swahili Dialects of Somalia and the Northern Kenya Coast." In *Etudes sur le bantu oriental: Comores, Tanzanie, Somalie et Kenya*. LACITO documents Afrique 9. Paris: SELAF.
- Shariff, Ibrahim Noor. 1988. *Tungo Zetu*, Trenton, New Jersey: Red Sea Press.
- UNESCO. 1995. *UNESCO Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social and Human Science, Culture, Communication and Information*. Paris: UNESCO Publishing.
- Werner, Alice. 1927. "A Traditional Poem Attributed to Liongo Fumo." *Festschrift Meinhof* (editor anonymous), pp. 45-54. Hamburg: L. Friederichson.
- Yahya Omar and Kevin Donnelly. 1982. "Structure and Association in Bajuni Fishing Songs." *Genres, Forms, Meanings: Essays in African Oral Literature*, ed. by Veronika Gorog-Karady, pp. 109-122. Oxford: JASO Occasional Papers (No. 1).
- Yahya Ali Omar and Peter Frankl. 1994. "Swahili Letters from the Taylor Papers (being a study of SOAS MS 47769 and MS 47782)" *South African Journal of African Languages* 14, supplement 1.

