This is a repository copy of *Modern South Arabian: Conducting Field work in Dhofar, Mahrah and Eastern Saudi Arabia*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/151587/

Version: Accepted Version

Book Section:

Modern South Arabian:
Conducting Fieldwork in Dhofar, Mahrah and Eastern Saudi Arabia[1]

Janet C.E. Watson, Miranda J. Morris, Abdullah al-Mahri,

Munira al-Azraqi, Saeed al-Mahri and Ali al-Mahri

Abstract

In this paper, we discuss conducting community-based fieldwork with speakers of the Modern South Arabian languages (MSAL) in southern Oman, eastern Yemen and eastern Saudi Arabia for a Leverhulme-funded project: *The Documentation and Ethnolinguistic Analysis of Modern South Arabian.* The paper begins with a brief introduction to the languages, their varying degrees of language endangerment, and the traditional lifestyle of their speakers. In section 2 we discuss the decline and erosion of the languages, and the rationale this provides not only for documenting the languages, but also for closely involving native speakers and community members in the data collection, transcription, translation, analysis, and dissemination. This vital community participation is considered in section 3, which also includes a description of the equipment we used, the software packages and the orthography devised for the project. We describe the collection of audio, audio-visual and photographic material, file identification and metadata, identifying speakers, obtaining ethical consent, training community participants, analysing and archiving the data, and the project website. Section 4 discusses language revitalisation and the joint dissemination of research.

1. Introduction

The Modern South Arabian languages are six endangered Semitic languages spoken in the southern extremities of the Arabian Peninsula: eastern Yemen, southern Oman, Jiddat al-Ḥarāsīs, the island of Soḳoṭra and southern and eastern portions of Saudi Arabia. These are: Mehri, spoken over the largest area, spanning eastern Yemen, southern Oman and reaching into southern and eastern Saudi Arabia; Śḥerɛ̄t (also known by some as Jibbāli), spoken in the mountains and coastal regions of Dhofar; Hobyōt, spoken in a small area spanning the Yemen-Oman border; Ḥarsūsi, spoken in the Jiddat al-Ḥarāsīs in central Oman; Baṭḥari,

once spoken around the coast and in the plateau above in eastern Dhofar, opposite the Al-Ḥallāniyāt islands, but now with very few remaining speakers; and Soḳoṭri, spoken on the island of Soḳoṭra.

The languages vary in endangerment from critical to moderate. In terms of speaker numbers, they range from some 20 for Baṭḥari to over 180,000 potential speakers for Mehri: the precise number of speakers is impossible to determine with any accuracy as there are no census figures relating to MSAL speakers specifically, and many members of the language communities no longer speak the languages fluently or at all. However, Baṭḥari, Ḥarsūsi and Hobyōt have the fewest speakers and are the most critically endangered languages. In this paper, we do not broach the issue of what constitutes a 'dialect' or a 'language'.

Traditionally, the Mahrah and Ḥarāsīs were nomadic camel, goat and sheep herders, with the Mahrah along the coast being heavily involved in trade and the commercial fisheries; the mountain and coastal-based Śḥerɛ̄t speakers led a less nomadic existence, transhuming seasonally between caves or temporary dwellings in the mountains with their cows, goats and camels, fishing on the coast and cultivating cereals and legumes in enclosed plots in the mountains during the monsoon; the Baṭāḥirah lived predominantly from fishing but also kept a few goats, sheep and some baggage camels; the Soḳoṭrans herded goats, sheep and cattle, and raised a few camels and donkeys as baggage animals; they fished along the coast and cultivated date-palms and finger millet where water resources allowed; Hobyōt speakers fished along the coast, and herded their cows, goats and camels in the mountains, and in the monsoon months cultivated enclosed plots to produce cereals and legumes for local consumption and for barter.

Multilingualism, at least in terms of comprehension, has always been the norm for MSAL speakers who are in regular contact. For example, in areas where they live close together, it is common for Mehri speakers to speak Mehri, and Śḥerɛ̄t speakers to respond in Śḥerɛ̄t and for both to understand the other perfectly adequately. The same would not be true, however, for a Śḥerɛ̄t and a Baṭḥari speaker, for instance, and none of the mainland MSAL speakers speak or understand Soḳoṭri.

2. Endangerment factors and language erosion

The MSAL languages lack any traditional script, which means that any script-based education or communication is conducted almost exclusively through Arabic. The lack of a

traditional script has two effects: (i) it results in a lack of written records for the MSAL, and (ii) it endows the languages with low prestige in relation to written languages. Since the word *luġah* 'language' in Arabic is used generally to refer to a written language, the MSAL are frequently described by the Arabic word *lahajāt* 'dialects'. The spread of Arabic in southern Oman from the 1970s and in Yemen and Saudi Arabia from earlier times has meant that the MSAL have been subject to steady erosion. This does not mean necessarily that fewer community members speak the languages, but that each subsequent generation has a reduced knowledge of the languages. Thus, younger speakers may describe themselves as fluent speakers of Mehri or Śḥerɛ̄t, but they do not have the command of, in particular, the vocabulary and morphology of their parents and grandparents. We present a few examples below.

Numbers: in southern Oman and eastern Saudi Arabia most speakers of Mehri or Śḥerɛ̄t over the age of c. 30 can count up to 30 and above. Younger speakers count up to 10 in their language, but from 11 onwards frequently use Arabic. When Watson was learning Śḥerɛ̄t in 2013, she was taught to count up to 30, but was told by schoolchildren in the Śḥerɛ̄t-speaking household in which she was staying that they were unable to count above 10 in Śḥerɛ̄t. In eastern Yemen, few speakers of any age regularly count above 10 in Mehri or Hobyōt, with some Mahriyōt-speaking children coming up with imaginative ways of forming numbers above 10, for example by using forms such as *ṭād ṭād* 'one-one' to express 11, and *ṭād ṯroh* 'one-two' to express 12.

Colours: the MSAL have a small set of basic colour terms for 'red', 'black', 'white' and 'grue' (which covers the 'yellow-green-blue' spectrum). In Mehri, these are: (ʕ)*ōfar* 'red', *ḥōwar* 'black', *albōn ~ ūbōn* 'white' and *ḥṣawr* 'grue'. To describe the colours of livestock the inventory of colour terms increases significantly. These terms are known and used by adults, but children—even children of goat and camel herding families—frequently fail to recognise these, and tend to resort to Arabic terms for the more basic colours: for example, using *aḥmar* 'red' in place of (ʕ)*ōfar* 'red' to describe a bay or chestnut camel, cow or goat.

Discourse markers: discourse markers are frequently replaced by Arabic expressions among all ages of speakers. Thus we observe Arabic *maṯalan* 'for example' in place of Śḥerɛ̄t *gens* and Mehri *hīs* or *hōbah*; *lākin* 'but' in place of Śḥerɛ̄t *duʰn* and Mehri *lahinnah*; *tamām* 'fine' in place of Śḥerɛ̄t *ḥayśōf* and Mehri *histaww* or *xayban.*

Nor is this loss of vocabulary being replaced with new MSAL terms to describe the modern world and its artefacts and technology. Instead speakers use Arabic, or an

arabicised version of the original in English, Hindi, Urdu and so on. All the MSAL are threatened by intense and rapid social, environmental and economic change, and the collapse of traditional cultural activities which accompanies such change. In Saudi Arabia, many of the Mahrah have left their traditional homelands in the desert interior and moved to cities such as Dammam and Al-Ahsa. Within Dhofar, until the 1970s there were no state-run schools or hospitals, transport was by foot or baggage animal, and water was collected on foot from natural sources. People lived in caves, or in brushwood or stone huts which they constructed themselves, and made use of tools and equipment which they made from locally sourced materials. Today the region enjoys all the trappings of the modern age. Younger generations no longer have, require or understand the extensive knowledge and practical skills of their elders and much earlier expertise has been lost or is disregarded, with imported alternatives replacing locally manufactured items. Traditional methods of natural resource and water management are no longer passed down to the next generation, and significant degradation of the environment has been the result. There is severe overgrazing in some areas, mismanagement of the increasingly scarce water supplies, and the sea is seriously overfished. This means that plants and animals that once played a significant role in everyday life are now extinct or rare. Language and culture are intrinsically linked, and this loss of traditional knowledge, skills and habitat is one of the key factors in language endangerment across the globe, particularly, but not exclusively, in the lexis.

Even where the languages continue to be spoken, the significance behind many figurative terms may be lost through urbanisation and modernisation: The positive connotations of Mehri *rwāġāb* 'branch [diminutive]', used to describe a young woman with height, slenderness and freshness, of *gizyɛrōt* 'little old milking animal' to describe a woman affectionately, and of *ḳayṣar* 'leopard' to describe a man of renowned bravery and strength may be lost in a society which no longer relates closely to the environment and livestock.

3. Community-based fieldwork

In a push to document and revive interest in the languages and cultures of Dhofar and Yemen, a Leverhulme-funded project (2013–2016) led by Janet Watson with Miranda Morris and Domenyk Eades as co-investigators, Saeed al-Mahri as principal local researcher, and Hammad al-Soqotri as administrative assistant, has been conducting audio, audio-

visual, photographic and textual documentation of the MSAL of Oman and mainland Yemen. Since 2012, Munira al-Azraqi has complimented our study with her research on the Mehri spoken in eastern Saudi Arabia. We have also recently produced culture-specific e-books in the languages for Mehri- and Śḥerēt-speaking children. We are very aware that it is only with the direct contribution and interest of community members that any plans for documenting and revitalising the MSAL can have any hope of success. The academic investigators see themselves primarily as a catalyst, stimulating interest and demonstrating how such documentation can be achieved. The intention is that thereafter their direct involvement will decrease as community members become more involved. To date the project has recruited over 100 speakers, several data collectors and data interpreters, and community members to transcribe the data and translate it from MSAL into Arabic. We believe that our project could be used as a template for future projects throughout the Gulf region and beyond to document language varieties and the vanishing cultural and environmental lexis.

3.1 Selection of equipment

Equipment was selected on the basis of sound quality, capacity, robustness and ease of use. For all UK investigators and local data collectors, we selected the Olympus LS11, a telephone-sized recorder that does not need an external microphone, and that can record in optimum loss-less format (PCM 44100 Hz, 16-bit stereo). It also has buttons large and clear enough to be used by our elderly consultants. A significant advantage of this device over the larger, more expensive Marantz is its economy in batteries: while the Marantz can use up four AA batteries within half an hour of recording, the Olympus LS11 needs only two AA batteries, and these last for several hours of recordings. This reduced the workload for the local research and administrative assistants in that they needed to check the batteries of data collectors' recorders every few weeks rather than on a daily basis. However, we also took out two Marantz PMD661 solid-state recorders, one for use in Dhofar and the other for use in Jiddat al-Ḥarāsīs. The Marantz PMD661 requires an external microphone and is best with external power, but has a significant advantage of sound quality over the Olympus LS11.

Audio-visual recordings were made by Watson and Eades to illustrate descriptions of cultural artefacts and to record the role of gesture in language. For recording gesture, these

were taken indoors, but either indoors or outdoors for illustrating cultural artefacts and activities. A selection of video recorders was used: the hand-held Everio JVC camcorder, which had the advantage of cost, size, weight and ease of use, and later a professional-level video recorder, the Canon XA20 with an external Audio-Technica microphone and a robust tripod. Alongside audio-visual recordings in the field, specific task-related audio-visual data was collected by Jack Wilson and Janet Watson from four MSAL consultants in the University of Leeds phonetics laboratory, using set map and shape tasks (Watson and Wilson 2017).

We collected two types of instrumental phonetic data from Mehri and Śḥerɛ̄t speakers, principally at the University of Leeds. The Laryngograph with Speech Studio software was used to examine laryngeal activity. The Laryngograph provides information on vocal fold contact. The technique is non-invasive and straightforward: two electrodes are placed on the throat positioned either side of the thyroid cartilage and a weak constant voltage is passed from one electrode to the other. This allows the current to fluctuate depending on variations in contact between the vocal folds: when the glottis is closed there is maximal conductance. One advantage of this equipment is that it is highly portable and easy to use, which meant we could take it into the field. Al-Azraqi, Naïm, Heselwood and Watson used the Laryngograph prior to the current project to examine vocal fold activity in the articulation of emphatic laterals in south-west Saudi Arabia (Heselwood et al 2012, 2013). The disadvantage of the Laryngograph is its price—upwards of £6000. However, once purchased, the software can be transferred to an indefinite number of computers. Since Johnstone in the 1970s (Johnstone 1987, etc.), researchers of MSAL have noted the patterning of emphatic and canonically voiced consonants to the exclusion of aspirated consonants; the Laryngograph has enabled us to investigate the degree to which this patterning is reflected in vocal fold contact patterns.

Electropalatography (EPG) with Articulate Assistant software was used at the University of Leeds to measure tongue-palate contact. This was particularly important for examining the articulation of the many sibilants and lateral consonants of the MSAL, and of Śḥerɛ̄t in particular. EPG work requires individual palates with sensors to be produced for each speaker recorded. However, due to the size and value of the hardware, such data needed to be collected at the University of Leeds, which meant that, to date, EPG data has only been collected from five individuals for this particular project.

3.2 Orthography and selection of software

At the beginning of the Leverhulme project, the UK investigators devised an Arabic-based orthography for use by community members. The new orthographic characters were selected on the basis of Arabic symbols commonly adopted by native speakers in texting in MSAL. In early 2010 Watson had begun texting in MSAL with native speakers, using Arabic symbols. Since Śḥerēt exhibits more consonants than the other mainland MSAL, the base orthography was based on Śḥerēt, with the other languages utilising a subset of the Śḥerēt characters. In contrast to the Russian team led by Naumkin (cf. Naumkin, Kogan et al 2015), we did not add vowels to the new orthography. Our rationale for this decision was based firstly on the fact that context allows vowel quality to be inferred, and secondly on our need for a system that was maximally simple to use: the more complicated the system, the less likely would speakers be to adopt it. In developing the system, we made maximum use of Arabic characters, even where sounds do not have the same reflex as in most Arabic dialects. This is the case with the emphatics, with many tokens being realised as ejective (Ridouane et al 2015): thus, Arabic ض was adopted for the lateral emphatic, as cognate of ض; ظ for the interdental emphatic; ص for the alveolar sibilant emphatic; ق for the velar emphatic; and ط for the alveolar plosive emphatic. The emphatic alveo-palatal sibilant, /ṣ̌/, was represented by ص with three superscript dots, thus: ڞ. This symbol was selected on the basis of ض being the emphatic counterpart of ش, just as ص is the emphatic counterpart of س. The voiceless lateral fricative, /ś/, was transcribed as an inverted ث, thus: پ. This choice was made on the basis of native speakers most frequently adopting ث for this phoneme in texting, and because the other possible base symbol, ش, was required in Śḥerēt to cover both the alveolar-palatal fricative, /š/, and the rounded, hushing-hissing sibilant, /s̃/ (cf. Bellem and Watson, 2017). The Arabic ش was adopted to represent Śḥerēt /s̃/, since this phoneme corresponds to /š/ in the other mainland MSAL, and the /š/—less frequent in Śḥerēt—was represented by ش with three subscript dots, thus: ڜ. Nasalisation in Śḥerēt was denoted by a superscript ن, as in: ’̈. Finally the voiced lateral fricative allophone of /l/ in Śḥerēt, [ź], was represented by ذ with two dots, thus: ‍ذ. This choice was made because Śḥerēt speakers most frequently used ذ to denote [ź] in text messaging. Saeed al-Mahri recruited and trained community members in the use of this new orthography, installed Times New Roman on their computers, and provided easily remembered short-cuts to the new symbols.

We then selected two freely available pieces of software for use by community members and investigators. PRAAT: Doing Phonetics by Computer: http://www.fon.hum.uva.nl/praat/, and later ELAN annotation software: https://tla.mpi.nl/tools/tla-tools/elan/. PRAAT enables transcribers to see the sound waves and spectrogram in transcribing, and to listen to the same section multiple times, but transcribers of longer texts have to transcribe into Word. ELAN is a more complicated programme, requiring the insertion and labelling of default and annotation tiers, but has the advantage of directly aligning the transcription and the English and Arabic translations with the media. By year two of the project, ELAN files with tiers for transcription, English translation and Arabic translation were being produced. These files were then sent together with the relevant sound or video files via Dropbox to translators, who were trained to click on the Arabic translation tier, listen to the relevant section of the recording, and then insert the translation. Completed ELAN files are archived with Endangered Languages Archive (ELAR) http://elar.soas.ac.uk/ at SOAS, University of London, together with the sound/video files.

One of the intended project outputs are cultural glossaries of the six MSAL. For this, we selected Toolbox software: http://www-01.sil.org/computing/toolbox/, a package developed by SIL International. With considerable help from the programme developers in the United States, Watson produced a template to be used by all investigators, including set field markers, semantic fields and parts of speech. We have also produced a comparative basic cultural glossary across all six of the Modern South Arabian languages, which is currently in press as a Journal of Semitic Studies supplement (Morris, Watson with Eades, in press).

3.3 Ethical consent and identification of speakers

Ethical consent forms were produced in English and translated into Arabic for speakers and data collectors. Due to the sensitivity official documents can provoke, it was decided early in the project to explain the project aims verbally instead, and to obtain ethical consent orally. Within the speaker/participant metadata sheet, a note was made as to whether a participant wished to remain anonymous in all outputs, and the degree of access they would permit for their materials. Women tended to prefer to remain totally anonymous, although a few opted for the Arabic-style Umm Abdullah, Umm Ahmad, etc., and several who agreed

to be recorded gave permission for their sound files to be archived, but requested that these sound files not be made accessible to people outside their local communities (as on YouTube or community websites).

We identified speakers through personal contacts and word-of-mouth, and later with assistance from data collectors. The difficulty in recruiting speakers is often ignored in scientific reports, but may be one of the most stressful aspects of fieldwork. As Saeed al-Mahri says, 'some people had doubts about the real purpose of the project and related it to the colonial periods, but the majority appreciated the initiative and the efforts made by the researchers.' We also had the experience of people who were unwilling to hear their own voices, or who doubted their knowledge, or who feared any recording would be shared with the world on YouTube.

Where personal contacts do not yet exist, the recruitment of speakers is even more of a challenge. In this regard, Munira al-Azraqi had an interesting experience. She had learnt about Mehri through contact with Watson and Ali al-Mahri, and realised through a YouTube video that Mehri was spoken near to her area of Dammam in Saudi Arabia. She says: 'It was difficult to find Mehri speakers. I started looking in the Souk in Dammam city. I looked in many places but couldn't find anyone at first. Then I found that people working in a shopping centre were using Arabic, but had an accent I didn't recognise. I knew that they were Mahrah. I asked one of them; he didn't answer my question, but looked at me and said, 'What do you want?' I told him I just wanted to learn about the language. He ignored me and kept on walking, but I followed him, repeating my request and telling him my name and where I work. He called one of his colleagues over, and then three gathered around me. I showed them my business card. I told them that I realised it was not appropriate for me to talk to them like this, but that I needed them to help me to find a woman to teach me Mehri because I wanted to tell my students about it. One of them took my card and agreed to give me his number. A few days later, a lady called me and agreed to teach me. The speakers didn't appear very interested to share this language with a stranger. We started to meet in the Corniche in Dammam every Friday for about four months. It wasn't easy to do the interviews because she didn't allow me to do any recording. I started to look for other informants to check the data I got and to get some recordings. Eventually I found a university student whose mother language was Mehri. She said she was born in the desert interior and raised in Dammam. I worked with her and her family and later with other men.

The Mahrah are very well connected with each other, and seem to instinctively know where you are and what you are doing. Once they get to know you, it is easy to work with them'.

When working with a large number of people, investigators soon come to recognise the particular strengths of individuals. Watson's first Mehri consultant, Askari Hugayran, was meticulous in terms of timekeeping, and enjoyed nothing more than to go through verb and preposition paradigms; Khalid Ruweya al-Mahri displayed the patience of a saint in teaching pronunciation and editing transcriptions; Ali al-Mahri proved to be an excellent co-presenter at conferences and invited lectures; Abdullah Musallam al-Mahri and Ali al-Mahri both showed an instinctive understanding of phonetics; Saeed al-Mahri has a particular interest in the use and collection of figurative language; and Abdullah al-Mahri and Ahmad Hardan display excellent interviewing techniques in data collection.

Morris and Watson, as female investigators, were able to work with both male and female speakers; however our male colleague, Domenyk Eades, working in the Jiddat al-Ḥarāsīs, was unable to work with female speakers. Luckily we had some hundred recordings collected by Morris in the 1970s and 1980s of women speaking Ḥarsūsi, and during the project a number were collected from one of the same women. The Baṭḥari data collector, Khalifa al-Bathari, became a very enthusiastic member of our team. Although at the start of the project he found using the recording device troublesome and was uncertain about how to go about collecting data—he tended to ask a speaker to 'just talk about anything from long ago'—after working with Morris on field trips he became more confident, and started to plan in advance what subjects he was going to raise and which stories he was going to ask for. He encouraged people to remember poems and to sing songs and he also devised differing data collecting techniques, such as sitting two speakers together and encouraging them to converse together on a chosen subject. Although his questioning and his comments and interjections are largely in the local dialect of Arabic on the recordings, his use of Baṭḥari, not spoken by him since late adolescence, improved significantly over the course of the project.

Working with speakers of languages such as Baṭḥari and Hobyōt where there is severely limited choice of speakers is not without its challenges. These range from interpreting sounds issuing from behind the thick cotton, face-covering mask formerly worn by the older women of the Baṭḥari community, to impaired hearing in speakers resulting in very loud speech, or the inevitable loss of teeth in the elderly sometimes making speech unintelligible. This loss of teeth makes differentiating between the sibilants particularly

tricky, though it is remarkable how repeated listening to the same voice can accustom the listener to a speaker's idiolect. Most marked though was the inability of the younger members of the community to understand the vocabulary used by elderly speakers. This meant that data analysis had to be done in the homes of speakers, imposing further on their space and time, and with all the distraction, interruptions and background noise that this entails. An alternative was analysis in Salalah on a one-to-one basis with a younger, less fluent speaker, but this necessitated frequent phone calls to an elderly relative to check on a particular term or phrase, or making notes to ask speakers at a later date about certain procedures or practices. There were also other challenges posed by Baṭḥari speakers, largely consequent on their language having a low status. Three Baṭḥari speakers, two women and a man, were the children of a Ḥarsūsi-speaking mother. This was unusual in that low-status Baṭāḥirah were rarely permitted to marry higher status non-Baṭḥari women. But the Baṭḥari father of these three had been one of a handful of Baṭḥari traders who had managed to accumulate a large herd of goats and camels. He lived in the desert plateau behind the coast and here he had met and married a Ḥarsūsi woman. The influence of their mother's language (Ḥarsūsi is very close to Mehri) is occasionally noticeable in the speech of all three children, especially when reciting poetry or singing. Interestingly, the son, who also raises camels, the prestige animal of the region and one closely associated with the Mahrah and the Ḥarāsīs, demonstrated a pronounced Mehri 'twang' or flavour when he recorded material on camels and camel rearing. Another Baṭḥari speaker, orphaned at a very young age, had been helped and looked after by a man speaking Mehri and his Baṭḥari wife, and this is again occasionally evident in his speech (he sometimes uses the Mehri *hawḳaʔ,* 'to put, place' instead of the Baṭḥari *awḳaʕ*, for example). Luckily there are enough speakers of purely Baṭḥari parentage to make it possible to recognise such deviations. The main problem then was to try and disentangle the vocabulary and morphology of the local Arabic dialect which all Baṭḥari speakers use in their everyday life from the Baṭḥari.

3.4 File identification and metadata

Metadata is produced for participants and for each audio, audio-visual, textual and photographic file. Working with a large team that includes non-academic members, it was essential to provide a clear template for file labelling and metadata listing. Each file has an identification which includes: date, in the form YEARMONTHDAY; language name; dialect

code; speaker code; topic; file format. For example: *20140612_MehriRabkut_M021_movingtopasture.WAV* refers to a Mehri sound file in the Rabkut dialect recorded by speaker M021 on 12th June 2014 on the topic of moving livestock towards pasture, and saved in WAV lossless format (44,000 Hz, 16 Bit). *20140612_MehriRabkut_M021_movingtopasture.eaf* is the transcribed and translated ELAN file produced from this sound file.

Metadata for sound and video files includes date; language name; dialect code; speaker age, region, education level; topic details; place of recording; environment in which the recording was produced; any special features in terms of syntax, lexicon, file content; and extraneous factors that may have affected the recording, such as noise level.

3.5 Data collection and analysis checking with native speakers

The digital revolution has transformed the way in which data can be collected, checked and analysed. For the collection of data by local data collectors while the UK investigators were not in the field, we registered Dropbox accounts: https://www.dropbox.com/, and created folders which were shared with data collectors and the research assistant and the principal local researcher. Data collectors either brought their sound files on a regular basis to Saeed al-Mahri or Hammad al-Soqotri for uploading to Dropbox, or uploaded the sound files themselves. They then notified the relevant UK investigator by email. Saeed al-Mahri produced metadata for uploaded files in Mehri and Śḥerēt; Morris produced the metadata for uploaded files in Hobyōt and Baṭhari. Dropbox is also used for sharing sound and ELAN files with translators and translation editors in the field.

When we first began conducting fieldwork in the Arab world, analysis of data needed either to be conducted in the field or required us to bring language consultants to the UK for several weeks at a time. Through the ready availability of digital technology, several language consultants from among the younger Mehri and Śḥerēt speakers can be contacted in Arabic by email to check queries, and where email is not an option, the majority of our consultants have smart phones and use the voice and writing facilities of WhatsApp. When encountering words the investigators do not understand during transcription, they frequently send text or WhatsApp messages to consultants for clarification. For example, in the transcription of a text dealing with the camel udder harness, Watson sent a text to Abdullah al-Mahri to ask what يرطون meant and what the

singular form was. He replied بينيذ نزول بيريط طاد وطاد خييط = برطوتن '*śarṭawtan* means strings and one is *śīrēṭ* on the pattern of *ḏīnēb*'. All transcribed Mehri and Śḥerēt texts are sent with the sound/video file via Dropbox to Abdullah al-Mahri and Saeed al-Mahri respectively with queried sections highlighted between *. They respond by typing in the correct word or phrase using the new Arabic-based orthography: for example, *yiʔamram hēh *amaxsīs** امخسيس 'they call it *amaxsīs*', and **yikaśmam** يكِمم 'they stop [tapping the frankincense trees]'. Unfortunately this was not possible for Baṭḥari or Hobyōt data analysis, where speakers were illiterate and the younger literate members of the next generation had little or no knowledge of the language of their parents, so analysis had to be carried out on field trips or by bringing a data interpreter to the UK.

3.6 Archive production with ELAR

Since the onset of the project, we have collaborated closely with ELAR at SOAS, University of London, to archive all audio, audio-visual and photographic files. Metadata for these files was initially produced using an excel spreadsheet, but has since been changed to CMDI Maker, an on-line metadata resource produced in Cologne: http://class.uni-koeln.de/cmdi_maker/. CMDI Maker allows archivers to bundle files together—for example, where a video and sound file have been produced simultaneously, or where photographic material is available to illustrate the subject of a sound/video file. The archive includes all primary data files and corresponding ELAN files. Initially the team archived only ELAN files that had been doubly checked with speakers. Today we archive all files, noting through CMDI Maker whether the ELAN file is in final or draft form. Today five archives are available at ELAR (Baṭḥari: Morris 2016a: https://elar.soas.ac.uk/Collection/MPI984105; Hobyot: Morris 2016b: https://elar.soas.ac.uk/Collection/MPI985006; Mehri: Watson & Morris 2016a: https://elar.soas.ac.uk/Collection/MPI976775; Śḥerēt: Watson & Morris 2016b: https://elar.soas.ac.uk/Collection/MPI972274; Harsusi: Eades & Morris 2016: https://elar.soas.ac.uk/Collection/MPI984354). These archives contain hundreds of sound files with metadata (and in the case of Baṭḥari, thousands!), and audio-visual files and photographic stills in the case of the Mehri, Śḥerēt and Harsusi archives. A number of sound and audio-visual files are accompanied by ELAN transcription and translation files. We are currently working on transcribing more of the files.

3.7 Website

The website for the Documentation and Ethnolinguistic Analysis of Modern South Arabian, http://www.leeds.ac.uk/arts/info/125219/modern_south_arabian_languages, was produced towards the end of the first year of the project, with assistance from Leendert Plug. It is maintained by Watson with help from Rebecca Wilding at the University of Leeds: It contains details of all local and international participants, the scientific committee, and sections on news, events and resources. Since this is the main face of the project to the community and the outside world, we ensure that it is regularly updated and includes news about community member activities. The resources section includes a regularly updated bibliography of the MSAL, YouTube links, theses and dissertations on MSAL by Omani and Yemeni scholars, the first children's e-book in Mehri, photographs (current and historical) and videos, through Flickr: https://www.flickr.com/. There are also sample audio files: the numbers from 1–20 in each language, and a selection of cultural audio files. In recent months, Abdullah al-Mahri has started to publish a cultural blog on the website in Mehri, using the new Arabic-based orthography.

4. Language revitalisation and joint dissemination

Our expressed hope has always been that the community will take renewed pride in their languages and culture, and will continue to document these long after the UK investigators have left the field. Throughout the project we have tried to promote language revitalisation by encouraging speakers to speak their language and also to write it, in the hopes that they will encourage their children to do the same, using their own language as well as Arabic. Newspaper articles were written and talks on the importance and uniqueness of the MSAL given at local schools. Data collectors were encouraged to discuss language loss and its implications within their language communities, and many informal discussions were held with groups from the language communities, as well as with school teachers and staff at Dhofar University—always with the aim of raising the profile and status of the languages, not only amongst the speakers themselves but also in the wider Arab community.

This attempt to raise the status of the languages has already had a marked effect on the small Baṭḥari community. A formerly disadvantaged people of low status, with feelings of shame about their former poverty and lowly position, they had been quick to embrace

Arabic and the new ways of life. Morris had worked with the Baṭāḥirah in the 1970s and 1980s so was fortunate in having many personal contacts. At the start of the project, she held meetings with tribal elders to explain the aims of the project, which was met with some incredulity. Those few who still speak Baṭḥari are well known within the community, so it was not difficult to identify speakers, but they are all elderly and illiterate men and women, and were not enthusiastic initially about being singled out for this purpose. However they were too kind and embarrassed to refuse to help, and work began. The early recordings were stiff and lacked fluency, and their Baṭḥari was heavily adulterated with Arabic. However, over the period of the project, both language proficiency and attitude has changed markedly. Speakers began to talk with a new pride and confidence about the skills they used to support themselves and their families in the hard earlier times, and they began to recall songs and poems they used to sing and stories they used to tell each other. Increasingly younger members of the community chose to be present at recording sessions, their interest in learning about their past and their amazement at and respect for the ingenuity and survival skills of their forebears steadily growing. The enthusiasm of the Baṭḥari data collector for his language continues to grow, and to date, and including those recorded by Morris in the 1970s and 1980s, over 2,500 sound files of ever increasing fluency and interest have been recorded. The new curiosity of the younger members of the community about their former language was demonstrated on Morris' last fieldtrip when she was repeatedly passed a phone so she could listen to recordings of songs which she herself had made in the 1970s, and which, at his request, had been copied to the smartphone of Khalīfa al-Baṭḥari when he came to work with her in the UK. However, since none of the younger generation speak or understand Baṭḥari, it seems more than likely that this language will be lost within the next generation or two.

Hobyōt is not as endangered as Baṭḥari, but among the younger generation it has not been possible to find anyone who speaks Hobyōt as their primary language; they speak instead Śḥerēt or Mehri / Mahriyōt, and all speak, read and write Arabic. However, it seems that the project might have succeeded in making this generation more aware of the dangers of losing their language altogether. For example, one of the data interpreters was struck by how much of the material he was working on with Morris was completely new to him. He then began to discuss with Morris the possibility of embarking on a project to record his elderly mother, a Hobyōt-speaker from the extreme west of Dhofar. He has now bought a digital recorder and worked with Morris to draw up a list of subjects to explore with her,

with the aim of capturing and documenting the way of life of the past, and especially the expertise of women. To date, he has sent Morris five blocks of recordings made with his mother, for whom Hobyōt is her first language, and he is enthusiastic about making further recordings.

We have tried to foster a sense of community ownership through including local news and photographs of participants on the website, and encouraging shared dissemination of research in open lectures, workshops, webinars and publications. Since 2011, five native speakers of Mehri and Śḥerēt and one native speaker of Baṭḥari have come to the UK to assist in data collection and data analysis. Between 2011 and 2012, Mohammad al-Mahri took part in the first workshop on Mehri to be held at the University of Salford, co-presented with Domenyk Eades and Janet Watson at the Camel Conference at SOAS, London, and co-presented with Janet Watson at the Hajj Conference at the British Museum, London; Saeed al-Mahri participated in a phonetics workshop on Mehri in 2012, in a webinar on changing attitudes to water in Dhofar at the University of Leeds in 2014, and in a Mehri language workshop at Uppsala in 2018. Saeed is also on the water@leeds distribution list for which he produced a blogpost in 2015. Ali al-Mahri has co-presented with Janet Watson either in person or by phone on Mehri and Śḥerēt in workshops and conferences in Muscat, Dhofar, Doha, Erlangen, Frankfurt, Paris, Jedda, Leeds and London. Ali al-Mahri conducted Mehri language workshops in Salalah with Watson in January and August 2018. Khalid Ruweya al-Mahri came to the UK for two periods of one month in 2015, and participated in a workshop on endangered languages at the University of Leeds and co-presented a guest lecture at the University of Edinburgh. Abdullah Musallam al-Mahri came to the UK for six weeks in 2015 and for three weeks in 2018; he participated in a lecture on endangered languages, an international conference on Language and Nature in Southern and Eastern Arabia in Doha, a presentation on NLP for endangered Semitic languages in Doha, and in a videoed debate on Language and Nature at the University of Leeds. The video debate attracted 5,500 participants from over 100 countries.

Joint research publications have been produced with Abdullah Musallam al-Mahri (Watson and al-Mahri 2017, forthcoming), Salim Awadh al-Shahri (Morris and al-Shahri, 2017) and Mohammad al-Mahri (Eades et al 2013). One volume resulting from the project is currently in press (Morris, Watson with Eades, in press), and four volumes are currently in preparation with the collaboration of native speakers: a volume of Baṭḥari oral texts and a volume of Hobyot texts, both in preparation by Morris, and a pedagogical grammar of

Mehri, in preparation by Watson with Abdullah and Ali al-Mahri (all for publication with Harrassowitz), and the phonetics and phonology of Mehri and Śḥerēt, in preparation by Watson with Abdullah al-Mahri (for publication with OUP).

5. Conclusion

Collaborative fieldwork and data collection has its frustrations and challenges. Different attitudes to time-keeping, working hours and methods, and the differing priorities of team and community members have had to be taken into account, with the result that we have all had to make adjustments in various ways. Nevertheless, challenges can at times prompt imaginative solutions. For example, a lack of literate Baṭḥari speakers able to translate directly from Baṭḥari into Arabic has meant that for this language we now resort to translating the English translations into Arabic. Again, due to the difficulty in recruiting sufficient people to transcribe texts using the new Arabic-based orthography, Chris Norton from the University of Leeds worked with Watson to develop an automatic transliterator (Watson, Norton and al-Mahri 2018). This programme transforms material transcribed in the Latin-based transcription into the new Arabic-based orthography. Thus, only material which has not been transcribed in Latin-based transcription continues to be transcribed by hand on the part of native-speaker consultants.

A preparedness to make compromises with co-investigators and other team members is essential, as is the holding of regular meetings to discuss and revise project goals and outputs. Nevertheless, collaborative fieldwork can be extremely rewarding and instructive, and we believe that, over the long term, it can produce significantly better results than individually conducted fieldwork.

References

Bellem, Alex and Janet C.E. Watson. 2017. 'South Arabian sibilants and the Śḥerēt *s̃* ~ *š* contrast'. Laïla Nehmé and Ahmad Al-Jallad (eds), *To the Madbar and Back Again: Studies in the languages, archaeology, and cultures of Arabia dedicated to Michael C.A. Macdonald*. Leiden-Boston: Brill, 622–644.

Eades, Domenyk, Janet C.E. Watson, and Mohammed A. al-Mahri. 2013. 'Camel Culture and Camel Terminology Amongst the Omani Bedouin'. *Journal of Semitic Studies* 58 (1): 169–186.

Eades, Domenyk, and Miranda J. Morris. 2016. *The documentation and ethnolinguistic analysis of Modern South Arabian: Harsusi.* ID: Harsusi (0314). London: SOAS. Endangered Languages Archive, ELAR.

Heselwood, Barry C., Janet C.E. Watson, Munira Al-Azraqi, and Samia Naïm. 2013. 'Lateral Reflexes of Proto-Semitic *ḍ and *ḏ̣ in Al-Rubūʿah Dialect, South-West Saudi Arabia: Electropalatographic and Acoustic Evidence.' In Renaud Kuty, Ulrich Seeger and Shabo Talay (eds), *Nicht nur mit Engelszungen: Beiträge zur semitischen Dialektologie: Festschrift für Werner Arnold.* Wiesbaden: Harrassowitz Verlag, 135–144.

Johnstone, Thomas M. 1987/2006. *Mehri Lexicon and English-Mehri Word-List.* Oxford: Routledge.

Johnstone, Thomas M. 1981. *Jibbāli Lexicon.* London: Oxford University Press.

Morris, Miranda J. 2016a. *The documentation of Modern South Arabian languages: Bathari*. ID: Bathari (0364). London: SOAS, Endangered Languages Archive, ELAR.

Morris, Miranda J. 2016b. *The documentation and ethnolinguistic analysis of Modern South Arabian: Hobyot.* ID: Hobyot (0309). London: SOAS, Endangered Languages Archive, ELAR.

Morris, Miranda J., and Salim Awadh al-Shahri. 2017. 'Drink Long and Drink in Peace: Singing to Livestock at Water in Dhofar, Sultanate of Oman'. In Laïla Nehmé and Ahmad Al-Jallad (eds), *To the Madbar and Back Again: Studies in the languages, archaeology, and cultures of Arabia dedicated to Michael C.A. Macdonald.* Leiden-Boston: Brill, 610–621.

Morris, Miranda J., Janet C.E. Watson, with Domenyk Eades. Forthcoming. *A Comparative Cultural Glossary across the Modern South Arabian Language Family*. Journal of Semitic Studies supplement series, Oxford University Press.

Naumkin, Vitaly, Leonid Kogan, ʿIsa G. al-Daʿrhi, Ahmed ʿI. al-Daʿrhi, Dmitry Cherkashin, Maria Bulakh, and Ekaterina Vizirova. 2015. *Corpus of Soqotri Oral Literature. Vol. I.* Leiden-Boston: Brill.

Ridouane, Rachid, Cédric Gendrot, and Rajesh Khatiwada. 2015. 'Mehri Ejective Fricatives: An Acoustic Study'. Paper presented at the *18th International Congress of Phonetic Sciences*, University of Glasgow, 10 August 2015.

Tosey, Paul, Wendy Sullivan, and Margaret Meyer. 2013. 'Clean Sources: Six Metaphors a Minute'. http://www.cleanchange.co.uk/cleanlanguage/wp-

content/uploads/2014/02/Six-metaphors-a-minute-final.pdf (accessed 16 November 2015).

Traxler, Matthew J. 2012. *Introduction to Psycholinguistics: Understanding Language Science*. Chichester: Wiley-Blackwell.

Watson, Janet C.E., and Abdullah al-Mahri. 2017. 'Language and nature in Dhofar'. In Simone Bettega and Fabio Gasparini (eds), *RiCOGNIZIONI. Rivisti di Lingue e Letterature straniere e Culture modern*. University of Turin, 87–103.

Watson, Janet C.E., and Abdullah al-Mahri. forthcoming. 'A Stratal OT Account of Word Stress in Central Omani Mehri'. In Daniel Birnstiel and Na'ama Pat-El (eds), *Re-Engaging Comparative Semitic and Arabic Studies*. Wiesbaden: Harrassowitz Verlag.

Watson, Janet C.E., and Barry C. Heselwood. 2016. 'Phonation and Glottal States in Modern South Arabian and San'ani Arabic'. In Youssef A. Haddad and Eric Potsdam (eds), *Perspectives on Arabic Linguistics XXVIII*. Amsterdam-Philadelphia: John Benjamins, 3–37.

Watson, Janet C.E., and Miranda J. Morris 2016a. *Documentation of the Modern South Arabian Languages: Mehri*. ID: Mehri (0307). London: SOAS. Endangered Languages Archive, ELAR.

Watson, Janet C.E., and Miranda J. Morris. 2016b. *Documentation of the Modern South Arabian Languages: Shehret*. ID: Shehret (0308). London: SOAS. Endangered Languages Archive, ELAR.

Watson, Janet C.E., Chris Norton, and Abdullah al-Mahri. 2018. 'NLP for under-resourced Semitic languages: The case of Modern South Arabian'. Presentation at Qatar Computing Research Institute, Doha, 19th February 2018.

Watson, Janet C.E., and Jack Wilson. 2017. 'Gesture in Modern South Arabian Languages: Variation in Multimodal Constructions During Task-Based interaction'. *Brill's Journal of Afroasiatic Languages and Linguistics* 9 (1–2): 49–72.