

The project ‘Documentation and ethnolinguistic analysis of the Modern South Arabian languages’

This three-year community-based project funded by the Leverhulme Trust aims to document the Modern South Arabian languages (MSAL) spoken in Oman and mainland Yemen. The Modern South Arabian languages are Semitic languages spoken by minority populations in southern and eastern Yemen, western Oman and the fringes of southern Saudi Arabia. These languages belong to the South Semitic branch of the Semitic language family, which also includes Ethio-Semitic and Ancient South Arabian. This is distinguished from the Central Semitic branch, which includes the more widely known Arabic, Aramaic, and Hebrew.

The traditional culture of MSAL speakers is becoming increasingly fragile: urbanisation, rapid commercialisation, compulsory education in Arabic and a rise in living standards has resulted in the collapse of many aspects of the culture which used to characterise the region. This rapid economic and socio-political change has also resulted in the MSAL languages increasingly falling into disuse. The six languages of this group are in varying stages of endangerment: Mehri, spoken in Oman, Yemen and southern Saudi Arabia, is said to have some 100,000 speakers, although the actual number is difficult to estimate since the language is spoken across three state boundaries and many Mehri no longer speak Mehri; Soqoṭri, spoken on the island of Soqatra (and by a few hundred who have emigrated to the Gulf), has some 50,000 speakers; Šherēt, also known as Jibbāli, spoken within the Dhofar region of Oman, has some 20-30,000 speakers; Ḥarsūsi, spoken in Jiddat al-Ḥarāsīs of central Oman, and Hobyōt, spoken in the far east of Yemen and the far west of Oman, each have under 1,000 speakers; Baṭḥari, spoken in Dhofar, has fewer than 10 fluent speakers. Nowadays almost all speakers of MSAL also speak Arabic. This three-year project aims to document Baṭḥari, the MSAL about which least is currently known, Hobyōt, Ḥarsūsi and Šherēt; where necessary to collect additional culture-specific texts in Mehri; to produce a comparative 1,000-word glossary of culture-specific terms across all six MSAL; and prepare a comparative morpho-syntax of the six MSAL.

The MSAL are noted for their retention of ancient Semitic phonological and morphological features that have disappeared from other Semitic languages, and for innovations not attested in other branches of Semitic. They are unique within spoken Semitic for exhibiting a three-way number distinction (singular, dual, plural) in pronouns and verb inflections. Within the verb system, MSAL retain mood distinctions in the imperfect verb that have disappeared from other Semitic languages, and make widespread use of a continuous aspect marker with the perfect verb to indicate ingressive meaning, a feature unattested in other documented Semitic languages. All MSAL are unique amongst living Semitic languages in exhibiting three contrastive sibilants (s-like sounds) – *s*, *š* (sh) and *ś* (a lateral fricative hypothesised for Proto-Semitic). Furthermore, the lexis of MSAL shows links with Ethio-Semitic and with markedly conservative Arabic dialects of northern Yemen and south-western Saudi Arabia, suggesting early population movement and contact. In the early Islamic period, Arabic largely replaced the other languages of the Peninsula, and the nature of this loss and the impact the original languages have had on emerging varieties of Arabic is poorly understood in the literature. Thus the documentation and the synchronic and diachronic description of the MSAL is of crucial importance to understanding the historical development of the Semitic language family as a whole.

Alongside their importance to Semitic linguistics, the documentation of the MSAL is vital for recording the disappearing cultural traditions and socio-economic practices of the speakers. Language documentation provides an insight into the culture and way of life of peoples and preserves important traditional knowledge. This research will be of interest not only to linguists, but also to historians, anthropologists, sociologists, ecologists, zoologists and specialists in other scientific fields. Language documentation is also of great interest to the MSAL language communities who wish to maintain a record of their heritage and traditions in a world which is undergoing rapid change.

Over 700 Mehri texts have been published, dealing with narratives, folktales, oral histories, poetry, songs and cultural descriptions. For Soqoṭri, a body of some 1,000 songs and poems has been collected by Morris (Morris et al, in press) and oral texts have been collected by Morris and by the Russian mission. A handful of texts has been collected in Ḥarsūsi (Stroemer 2004). In order to produce an archive to enable comprehensive research on MSAL, it is necessary to collect further data for the least documented MSAL, in order of urgency: Baṭḥari, Ḥarsūsi, Šherēt and Hobyōt.

In collaboration with native-speaker researchers, between 15 and 20 hours of audio and some audio-visual, topic-focussed recordings will be collected in each language; around 40% of these will be selected for transcription, annotation and translation into both Arabic and English. To facilitate grammatical and lexical comparisons, the new texts will deal with subject matters covered by published sets of MSAL texts and will include: personal and local histories, particularly relating to the pre-1970s pre-motorised age, traditional livestock management, agricultural and fishing practices, milk-processing and food preparation, toponymy, pasture and water management, traditional skills and crafts, greeting procedures, date-harvest (for Ḥarsūsi and Baḥari), frankincense-harvest (for all but Baḥari and Ḥarsūsi), rhymes and games. The recordings of Baḥari, Šherēt, Mehri and Hobyōt made by Morris' 30–40 years ago will be compared with the project recordings, and we believe this will shed light on language change in the area. Photographic material will be collected to illustrate aspects of material culture, fauna and flora.

The transcribed, translated and annotated texts will be archived with the Endangered Languages Archive at the School of Oriental and African Studies, London: <http://www.hrelp.org/archive/> and prepared for book publication; a selection of the oral recordings will be made available on the project website; on the MahrahNet website, hosted in Dhofar, Oman; and on the Semitic Sound Archive, hosted in Heidelberg, Germany.

For the 1,000-term comparative cultural glossary terms will be selected from the text topics, from other culturally specific semantic fields and from lexemes identified in each of the MSAL; for those MSAL which exhibit considerable dialect variation, lexemes will be identified in each of the major dialect groups. The cultural glossary will be produced in Latin-based and Arabic-based transliteration, will be translated into both English and Arabic, and will be published on-line and in book form.

We very much hope that documentation of the MSAL and their cultural traditions and expertise will be continued by the language communities long after the period of our involvement.