

Background to the work

The six Modern South Arabian languages are spoken in Oman, Yemen, and parts of Saudi Arabia and the Gulf. This branch of the Semitic language family was first discovered by western researchers when the French consul, Fulgence Fresnel, wrote about Šheret in 1838. Since then work has been conducted on the MSAL primarily by the Viennese Expedition (late 19th – early 20th century), Thomas (1937), Johnstone (1975a,b, etc.), the French mission (from 1983), Sima (2009, etc.), Morris (1983, 1985, 2002, etc.), Liebhauer (2011), Rubin (2010, etc.), Watson (2009, 2011, 2012) and Watson & Bellem (2010, 2011). The last member of the family to be described to a western audience is Hobyot, identified by Johnstone in 1970s (Johnstone 1981).

Over 700 Mehri texts have been published, dealing with narratives, folktales, oral histories, poetry, songs and cultural descriptions (Hein 1909; Stroemer 1999; Sima 2009; Liebhauer 2011). A body of 1,132 songs and poems in Soqotri has been collected by Morris (Morris et al, in press) and the Russian mission (e.g. Naumkin et al 2011), and a handful of texts in Harsusi (Stroemer 2004). A body of oral cultural material has been collected for Baḥari, Šheret and Hobyot. In order to produce an archive to enable comprehensive research on MSAL, it is necessary to collect further data for the least documented MSAL: Baḥari, Harsusi, Šheret and Hobyot.

Lexical work on MSAL includes semantic-field-specific glossaries (e.g. Simeone-Senelle & Lonnet 1985–1986; Miller & Morris 1988, 2004), and bilingual dictionaries (Leslau 1938; Johnstone 1977, 1981, 1987). No detailed glossary of cultural terms has been drawn up for any MSAL, and comparative glossaries to date suffer from inadequate transcription and translation (e.g. Nakano 1986).

Descriptions of aspects of the grammar of MSAL include several articles (e.g. Johnstone 1975a, 1975b; Simeone-Senelle 2011; Lonnet 2009), early descriptions by the Viennese Expedition (Jahn 1905; Bittner 1909–14), grammars of Mehri by Jahn (1905), Wagner (1953), Rubin (2010), Watson (2012), and a syntax of Šheret (Hofstede 1998). The more that is known of the MSAL languages, however, the more evident it becomes that descriptions to date are inadequate, in many cases inaccurate, and lack sufficient insight into relationships between the languages.

Objectives

The MSAL are unwritten Semitic languages spoken by minority populations in southern and eastern Yemen, western Oman and the fringes of southern Saudi Arabia. The MSAL belong to the South Semitic branch of the Semitic language family, which also includes Ethiopian Semitic. This is distinguished from the Central Semitic branch, which includes the more widely known Arabic, Aramaic, and Hebrew.

The four objectives of this three-year interdisciplinary project are:

1. To compile a large corpus of annotated, transcribed and translated oral data from Baḥari, Hobyot, Harsusi and Šheret.
 2. To produce a comparative 1,000-term cultural glossary across the six MSAL;
 3. To collect and analyse acoustic and instrumental phonetic data of MSAL consonants;
 4. To establish parameters for a comparative grammar of the six MSAL.
1. Documentation of oral data will focus on topics covered by already collected sets of MSAL texts. The data set will include audio and audio-visual recordings of diverse oral traditions, including recitations of genealogies, folktales, poetry and songs. The number of people who can recite the oral traditions or indeed even passively understand them is dwindling rapidly. Recordings backed by still photographs will also include personal and local histories, particularly relating to the pre-1970s pre-motorised age, descriptions of traditional livestock management, agricultural and fishing practices, frankincense harvesting, food preparation, navigation by the stars,

traditional skills and crafts, greeting procedures, rhymes and games. The archive will incorporate digitalised recordings of previously collected material from the other MSAL, Mehri and Soqotri, and Morris's 1970s–1980s recordings of Baḥari and Hobyōt. The inclusion of material previously collected on Mehri and Soqotri will, for the first time, enable researchers to undertake typological research on this branch of the Semitic language family.

2. The comparative glossary will include the 1,000 terms deemed by local researchers and language consultants to be the most culture specific for the six MSAL languages. These terms will be drawn up with reference to Behnstedt & Woidich (2011), the text topics, any additional semantic fields (body parts, food/drink, cooking utensils, basic verbs of volition, movement, eating, bodily functions), and cultural terms identified separately in each MSAL; for Mehri and Šheret, which exhibit considerable dialect variation, lexemes will be identified in each major dialect group. Terms in the cultural glossary will be transcribed in IPA, translated into English and Arabic, annotated, illustrated with still photographs, and provided with links to electronic transcriptions.

3. Acoustic and instrumental phonetic data will be collected and analysed in order to establish the acoustic, articulatory and phonatory correlates of lateral fricatives and the third series emphatic consonants in the MSAL (see below).

4. Investigators will work with archive material to establish parameters for a comprehensive grammar of the six MSAL. Building on previous research conducted by Watson on Mehri, the most widely spoken MSAL, the comparative grammar will analyse key phonological, morphological and syntactic features of MSAL and produce a framework for further typological research on the languages of southern Arabia.

Hypotheses to be tested

1. The MSAL form a 'bridge' between Central Semitic (e.g. Arabic, Neo-Aramaic) and the Southern Semitic languages of east Africa (e.g. Amharic, Tigrinya, Tigre). Alongside phonological, morpho-syntactic and basic lexical features shared with both branches, MSAL exhibit both Central Semitic pharyngealised and Southern Semitic ejective reflexes of emphatic consonants.
2. The MSAL are the oldest continuously spoken Semitic languages. These languages feature: a) Three plain (non-emphatic) sibilants of which one is lateral. These have been reconstructed for proto-Semitic, but are no longer extant in other modern Semitic languages (Kogan 2011); b) Dual pronouns, no longer extant in other Semitic vernaculars; c) Mood distinctions exhibited stem-internally in the imperfect verb.
3. MSAL and the Ancient South Arabian languages are more closely related ethnolinguistically than has been previously hypothesised (cf. Huehnergard 2005). Lexical and textual project data will be examined in consultation with experts in ASA and Ethio-Semitic to identify linguistic and cultural links.

Significance

In recent decades, the spread of Arabic due to rapid economic and socio-political changes has resulted in the MSAL languages increasingly falling into disuse. The MSAL are in varying stages of endangerment: Mehri has between 100,000–180,000 speakers, although the actual number is difficult to estimate since the language is spoken across three state boundaries and many Mehris no longer speak Mehri; Soqotri, spoken exclusively on the island of Soqatra, has c. 50,000 speakers; estimates for numbers of speakers of Šheret, also known as Jibbāli, spoken in Dhofar in Oman, range between 10,000–30,000; Harsūsi, spoken in Jiddat al-Ḥarāsīs in Oman, and Hobyōt, spoken in far eastern Yemen and far western Oman, each have

under 1,000 speakers; Baṭḥari, spoken in Dhofar, has under 100 speakers. Nowadays almost all speakers of MSAL are also speakers of Arabic.

Alongside their importance to Semitic linguistics, the documentation of MSAL is crucial in documenting traditional knowledge of the disappearing ecosystem, cultural traditions, and socio-economic practices of the speakers. This documentation is of prime interest to the MSAL language communities themselves who are keen to record their heritage and traditions in a world undergoing rapid change. The project material will interest not only linguists, but also researchers in sociology, anthropology, ethnobiology, history, toponymy, ecology and other scientific fields.

This study will shed light on the dynamics of language change within a region where the related MSAL are undergoing loss at different rates. Baṭḥari and Ḥarsūsi are threatened by Arabic considerably more than Hobyōt and Šheret. The least prestigious MSAL, Hobyōt and Baṭḥari, are additionally threatened by Mehri. 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. The current documentation project will identify degrees of language decay in a region where Arabisation is currently taking place. Comparison of Morris's data from 1970s–1980s with 2012–2015 recordings will provide a historical perspective of language loss.

Features of typological linguistic interest

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.

MSAL are unique among extant Semitic languages in exhibiting a three-way distinction within the set of voiceless sibilants – /s/, /š/ and /ś/, a voiceless lateral sibilant similar to Welsh 'll' and hypothesised for Proto-Semitic (Kogan 2011).

Semitic languages are marked by exhibiting a three-way distinction in the alveolar and velar consonant series between voiced–voiceless–emphatic: in Arabic /d/ in *dall* 'to indicate' contrasts with /t/ in *tall* 'hill' and emphatic /ṭ/ in *ṭall* 'to look over'. Emphatics in Proto-Semitic have been hypothesised as having ejective realisation (Kogan 2011). Until 1970, Ethio-Semitic was believed to be the only extant Semitic language sub-family in which the emphatic consonants are realised as ejectives rather than as 'backed' or pharyngealised. Since Johnstone (1975a, 1975b), ejectives have been claimed to be a South Semitic feature, attested not only in Ethio-Semitic, but also in MSAL. The realisation of emphatics as backed in Arabic and Neo-Aramaic has been considered a Central Semitic innovation (Huehnergard 2005). However, recent work by Watson (2012) and Watson & Bellem (2010, 2011) suggests the move towards backed emphatics is not confined to Central Semitic. MSAL exhibit both ejective and backed emphatics, and in contrast to the ejectives of Ethio-Semitic all emphatics have a backing effect on adjacent vowels. Acoustic work in progress by Watson and Heselwood further shows that both ejective and backed emphatics exhibit the silent phase after release, typical of ejective consonants. This language family appears to be moving from an ejective to a backed emphatic system, thus the phonetic analysis of MSAL emphatics may shed light on the development of emphatics in Semitic.

The MSAL (excepting Soqotri) are the only Semitic languages in which the negative particle follows the negated term. Lucas (2009) and Watson (in press) suggest this word order is an innovation, and data available for MSAL indicate the languages occupy varying stages in Jespersen's Cycle in negation (Dahl 1979), an issue of typological interest for research on grammaticalisation and negation.

MSAL are unique within spoken Semitic for exhibiting a three-way number distinction in pronouns: singular, dual, plural. Other extant spoken forms of Semitic exhibit dual number only in nouns.

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.

The MSAL lexis suggests links between speakers of Ethio-Semitic and of conservative Arabic dialects of south-west Arabia, which exhibit ancient Semitic features not exhibited elsewhere in Arabic, including lateral reflexes of *ḏ, distinction between *ḏ and *ḏ̥, gender distinction in plural pronouns, and variable relative pronouns. The documentation and analysis of the MSAL will contribute significantly towards understanding both the historical development of the Semitic language family and population contact.

Methodology

Documentation will include audio and audio-visual records of culture-specific oral material in the field, still photographs, transcription of oral texts into IPA transcription and an Arabic-based script developed by the investigators with language consultants, morphological glossing, and translation of oral texts into Arabic and English.

Investigators and local researchers will collect 15–30 hours of audio recordings in each of Baṭḥari, Hobyōt, Ḥarsūsi and Šheret and 10 hours of audio-visual recordings. Audio data will include frame sentences for the acoustic investigation of consonants. Audio data will be collected on Marantz PDM661 recorders with dynamic microphones and Olympus LS-11 recorders. All audio material will be saved in WAV format 44.1 kHz, 16 bit. Audio-visual recordings will be made by investigators together with local researchers using Everio JVC camcorders. Audio-visual data will be saved in MPEG4 format. Still photographs will be taken using Canon PowerShot SX200IS. All data will be saved and stored on CDs/DVDs, 1000GB USB-powered hard drives by investigators and 32GB USB sticks by local researchers, and uploaded onto a 100GB Dropbox webportal. Subject to quality of recordings and ethical consent, local researchers will transcribe all data in Arabic-based transcription. 35–40 hours of audio data and 5 hours of audio-visual data will be selected for transcription into IPA, annotation and translation. ELAN software will be used to align transcription to media. 10% of the corpus will be morphemically glossed following Leipzig glossing conventions to enable morpho-syntactic analysis and inter-variety comparison.

Acoustic analysis using PRAAT software (<http://www.fon.hum.uva.nl/praat/>) will be used to establish the extent to which the so-called 'emphatic' consonants exhibit glottalic release. Instrumental phonetic data will be collected in the field using the Laryngograph (<http://www.laryngograph.com/>) to establish phonation patterns for emphatics. Further funds will be applied for to produce electropalatographic (EPG) palates for Hobyōt and Ḥarsūsi speakers to collect and analyse articulatory phonetic data at the University of Leeds. EPG data will be analysed to establish the articulatory features of the sibilants and the emphatic alveolar and palato-alveolar consonants.

The 1,000 glossary terms will be drawn up by investigators and local researchers from: the core concepts identified in Behnstedt & Woidich (2011); the corpus of transcribed texts; and region-specific field experience. The glossary will be organised according to semantic field, consonantal root, morphological pattern and word class. Flora and fauna items will be identified with reference to published works, including Gallagher & Woodcock (1980), Miller & Morris (1988, 2004) and Randall (1995), and further developed with local researchers and the fauna consultant, Alec Moore. Items relating to livestock husbandry and traditional skills and crafts will be selected by investigators and local researchers. All lexical items will be double checked with language consultants. The glossary will include links to audio-visual

footage and photographs taken in the field. For skills and crafts no longer practised, use will be made of museum and personal collections.

Parameters for the proposed comparative grammar will be identified on the basis of previous morpho-syntactic works on MSAL (e.g. Wagner 1953; Hofstede 1998; Watson 2012), with reference to general works on language documentation and description (e.g. Payne 1997), and works on language typology and specific aspects of grammar (e.g. Comrie 1976).

Details of how results will be published

Media products

Oral texts, annotations, transcriptions and translations will be held as a database with metadata records in XML format as prescribed by OLAC Metadata standards (<http://www.language-archives.org/OLAC/metadata.html>), archived for sustainability and accessibility on different archives, including: Semitic Sound Archive: (<http://www.semarch.uni-hd.de/index.php4>), and ELAR at Hans Rausing Foundation, SOAS: (<http://www.hrelp.org/archive>). They will also be stored in Salford Institutional Repository (USIR) (<https://usir.salford.ac.uk/>), and on the Mehri community website, Mahrah Net (<http://www.almahrah.net/vb/>).

Book and article publications

The transcribed, translated and annotated texts will be prepared for book publication. The glossary will be published on-line and in book form. The comparative grammar will be prepared as a book publication for Harrassowitz or Brill.

Articles on MSAL consonants will be targeted at journals specialising in phonetics, phonology and Semitic; articles on morpho-syntax at journals dealing with linguistic typology and Semitic; articles on oral literature, oral traditions and ethnobiology at the relevant specialist journals. Language community members will be helped and encouraged to present their own research on-line or in press in Arabic and using the Arabic-based script for MSAL developed for the project.