

A Guide to the Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic Language Turoyo

Mikael Oez

SOAS University of London

The main aim of language documentation is to create a long-lasting multipurpose record that captures the wealth of linguistic practices of a speech community. The purpose is to reflect traditions, customs, culture, civilization, etc. This article defines, navigates and provides insights into the contents of one particular language documentation project, namely the “Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic Language Turoyo”.

The documentation of Turoyo was funded by the Endangered Languages Documentation Programme (ELDP), at the School of Oriental and African Studies (SOAS), University of London. All the materials collected have been archived with the Endangered Languages Archive (ELAR), at SOAS, University of London. The materials held are digital and they are freely available to all users of ELAR at <https://elar.soas.ac.uk/Collection/MPI1035085>.

1. Introduction The digital age of technology has introduced a relatively novel approach to linguistics. The turn of the last decade of the 20th century attracted a new methodology of capturing language practices, which is known to us as “language documentation” (Himmelman 2006; Woodbury 2003). Technological advances enable the recording, processing, and storage of large amounts of high-quality linguistic data in portable devices. Fewer storage requirements provide more efficient codecs, thereby facilitating new perceptions and potential for fieldwork with language communities (Bouda et al. 2012).

The innovative methodology of collecting, analyzing, and decoding linguistic data is a stepping stone for enhancing and enabling interdisciplinary research in the humanities towards improving accountability for linguistic research results. Conventional outcomes of language documentation projects consisting of multimedia corpora, such as audio, video, photographs, and annotations, are interlinear text collections, dictionaries, sketch grammars and language revitalization materials.

Interest in linguistic diversity, and particularly in endangered languages, spreads beyond academia and has become a civic issue. Continuous reports on the subject have been published by the press and renowned institutions, such as UNESCO, with its *Atlas of the World’s Languages in Danger*. This activity has contributed to the increase of financial aid for the documentation and research into endangered languages. One of the bodies actively sponsoring the documentation of endangered languages is ELDP, from its headquarters at SOAS, University of London. The digitization of

languages has furthermore enhanced the need to standardize the documentation and study of endangered languages, and there is still an ongoing contention about how to adapt this methodology to other academic disciplines.

Language documentation archives use different software and interfaces to make linguistic data available. However, researchers use and find the most pragmatic and convenient way of storing their data according to their own collections, which makes data discovery and use, as well as comparative analysis across multiple corpora, extremely difficult (Woodbury 2014). It is therefore essential to provide a guide to the corpus to assist navigation through archived materials, with a description to uncover the history of the location, details of the consultants, or to explore a particular historical, socio-cultural, or grammatical topic. Each guide tells you how and where you can find, access, and understand the relevant records.

The Beth Qustan dialect project presented in this paper focuses specifically on old wives' tales and supernatural legends, resulting from the socio-religious influences of Islam and Christianity on one another. Such cultural stories include djinn stories told by native speakers, as well as episodes of visitations of Muslims and Christians to the cults of Christian saints, and consultation of soothsayers by Christians. Furthermore, aspects of daily life such as procedural texts, instructions, directions, and conversations are also included. It documents the socio-cultural practices of the Turoyo-speaking community mainly in Tur 'Abdin. The Beth Qustan dialect belongs to the Neo-Aramaic language known as Turoyo by its speakers, who are the Christian inhabitants mainly located in Midyat and in sporadic villages around it, which is in the central portion of the Mardin province in South Eastern Turkey. The area is referred to as Tur 'Abdin by its native speakers.

Following background information on the location, community, and the people who are the consultants of the project, this paper provides a bird's-eye view of the corpus of the Beth Qustan dialect, as well as insights into the contents. It further elucidates the methodology used for data collection and processing, with a focus on the project outcomes, and the essential imminent further research on the endangered Neo-Aramaic languages.

2. Background This section provides information about Turoyo speakers and their village with a focus on the language group presented in the Beth Qustan collection and the Turoyo language.

2.1 The Qusneans and Beth Qustan The Beth Qustan dialect of Neo-Aramaic is the vernacular of the Christian village of Beth Qustan or Beth Kustan. It is known in Neo-Aramaic as 'Be Qusyone', in Syriac 'Beth Qustan' (literally 'the house of Constantine'), in Kurdish 'Baqisyan', and it has become known by its Turkish name 'Alagöz' since the 1930s. Beth Qustan is one of dozens of villages in Tur 'Abdin (literally 'the mountain of worshipers') whose inhabitants speak a dialect of Neo-Aramaic. It is one of the few remaining Christian villages within the Christian stronghold of Tur 'Abdin, and it is a mountainous enclave of the Mardin Province, in South Eastern Turkey. Figure 1 indicates the location of the Beth Qustan village.

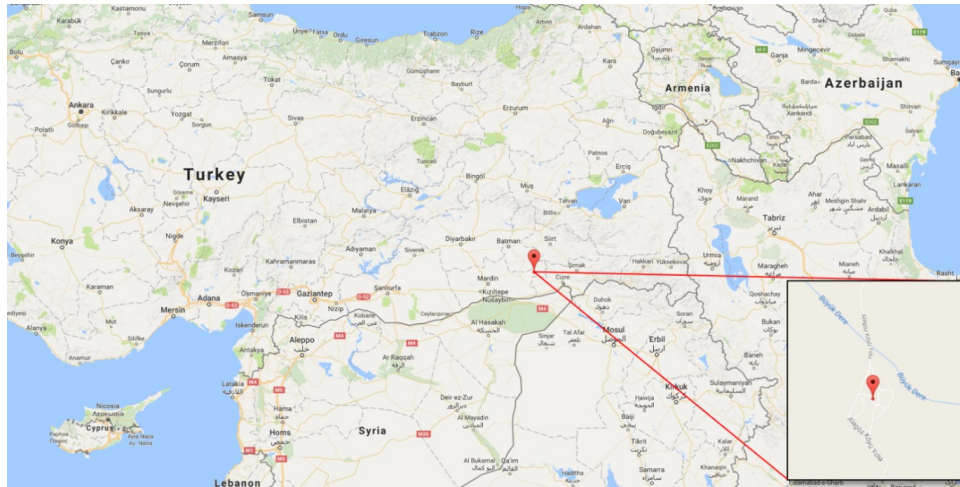


Figure 1. Location of the Beth Qustan village

The Beth Qustan village was a flourishing rural center, part of the food basket of the Fertile Crescent at the turn of the 20th century, with approximately 200 families living in the village; however, in 2017, only an estimated 20 families remained inhabitants of Beth Qustan. The Qusneans still speak a specific dialect of Neo-Aramaic, which is better known to the community as Turoyo, the language of Tur ‘Abdin. Figure 2 is a picture of the center of the Beth Qustan village.



Figure 2. Picture of the center of the village of Beth Qustan

Beth Qustan, like many other villages in the Tur ‘Abdin region, was probably inhabited in pre-Christian times, and it is believed that the village was a place of rest for travelers and that the army of the Roman Emperor Constantine I repeatedly marched through the village. Located near the village is the rock of Helen, which was named after St. Helen, the mother of Constantine I. Beth Qustan has nursed the renowned St. Gabriel, who was born in 574 and dedicated his life to the Monastery of Qartmin. Following the performance of several miracles, St. Gabriel became the patron saint of the Monastery of Qartmin, which by the end of the 15th century became known as St. Gabriel Monastery.

2.2 Group Represented The collection created within this project focuses on speakers of the Beth Qustan dialect of Neo-Aramaic of the Beth Qustan village of Tur ‘Abdin in South Eastern Turkey, who have immigrated and settled in diaspora. These are among the last remaining inheritors of the Aramaic language, which was the lingua franca of the Fertile Crescent from about 500 BC to the advent of Islam. It was the official language of the Achaemenid Persian dynasty in 559–330 BC. Whereas Aramaic continued to be used as the vernacular language in the whole of the Fertile Crescent, the official languages of the Roman and Persian Empires were Greek and Middle Persian, respectively. Nevertheless, the Aramaic language continued to be widely used until it was replaced by Arabic in ca. 650 AD.

All Aramaic dialects are now in danger of extinction, as most speakers of the Aramaic dialects have been forced, for many reasons, to abandon their original homelands and they have settled in diaspora over the last few decades. The younger generations of these communities are increasingly losing competence in these dialects due to their newly adopted homelands, cultures, and languages.

There are at least a couple of dozen Neo-Aramaic dialects that originated from Tur ‘Abdin which have yet to be documented and archived. Unless these dialects are documented prior to losing the current generation of native speakers, which will perish by 2025, we will find ourselves deprived of all invaluable knowledge on their culture and traditions. The Aramaic civilization goes back more than 3,000 years, but it has not been preserved, nor comprehensively documented.

The fieldwork collecting data for this project was initially due to be undertaken in Beth Qustan village in South Eastern Turkey. However, as a result of unforeseen political circumstances in Turkey on July 15, 2016, which had appalling consequences, a three-month state of emergency was declared on July 20, 2016. These unfortunate circumstances meant the cancellation of my planned travel in September–October 2016 to conduct fieldwork in the Beth Qustan village in Turkey. Particularly, my visit would most certainly be questionable if not untenable due to the sophisticated recording equipment I would consult in my fieldwork. Consequently, I had to consider the feasibility of a different location for my fieldwork. After thorough studies of the possibilities and logistics, I altered the location of my fieldwork and instead concentrated on native speakers of Beth Qustan in the diaspora community in Gütersloh and Kirchartd in Germany.

It is worth emphasizing that the generation I consulted for my fieldwork was born in the village of Beth-Qustan, and they still use their vernacular language in day-to-day communications, with minimum impact of the language of their new homeland. Fortunately, this sociolinguistic semi-isolation allowed me to remain within the remit of my planned project. Figure 3 is a photograph of one of the recording sessions with the diaspora community in Kirchardt, Germany.



Figure 3. (From left to right) Suleyman Agirman, Eliyo Acar, Lahdo Agirman and Iskender Demirel

A sociolinguistic overview of the Beth Qustan collection is illustrated in the following charts providing age group according to gender in Chart A, time of emigration from the homeland in Chart B, and the education of consultants in Chart C.

Chart A illustrates the age group according to gender, and it informs us that the majority of the consultants fall into the age group of 60–69. This is the age when consultants are still in their lucid years of remembering their childhood and can describe it with a wide range of vocabulary. This indicates that we have already lost the knowledge accumulated by octogenarians, as they found it much harder to remember things, and to talk in detail about their culture. The gender ratio of 3 to 18 is slightly disappointing, as females shy away from being recorded and the role of speaking in traditional gathering belongs to males. The females only agreed to being recorded collectively and in conjunction with traditional food cooking, which they felt more comfortable talking about, as house chores traditionally belong to their role as housewives.

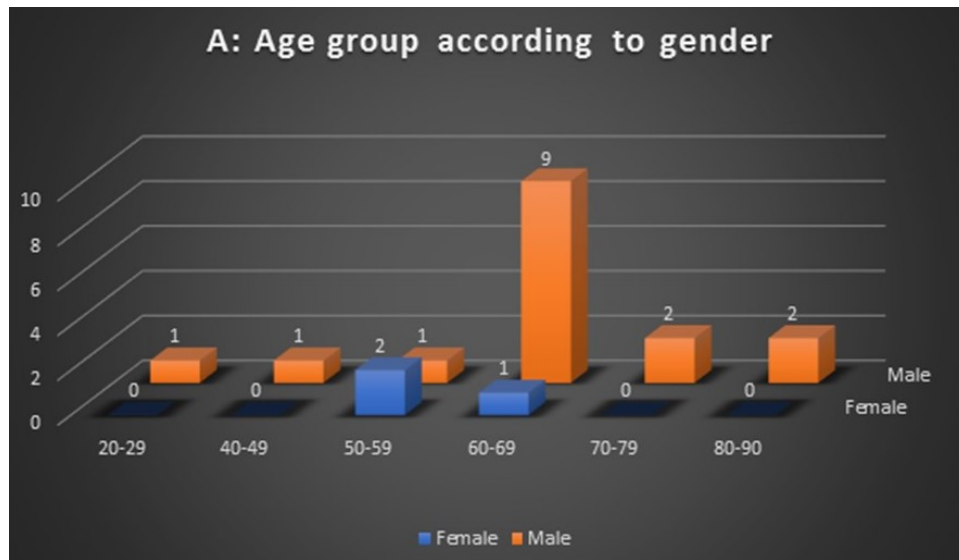


Chart A. Distribution of language consultants based on age group and gender

Chart B illustrates the decade of emigration of the consultants from the Beth Qustan village to Germany. Although there are several factors beyond the emigration, it is very clear that the peak of emigration was in the 1970s. This decade still preserved the old ways of living in the village, e.g., before electricity was introduced to the village.

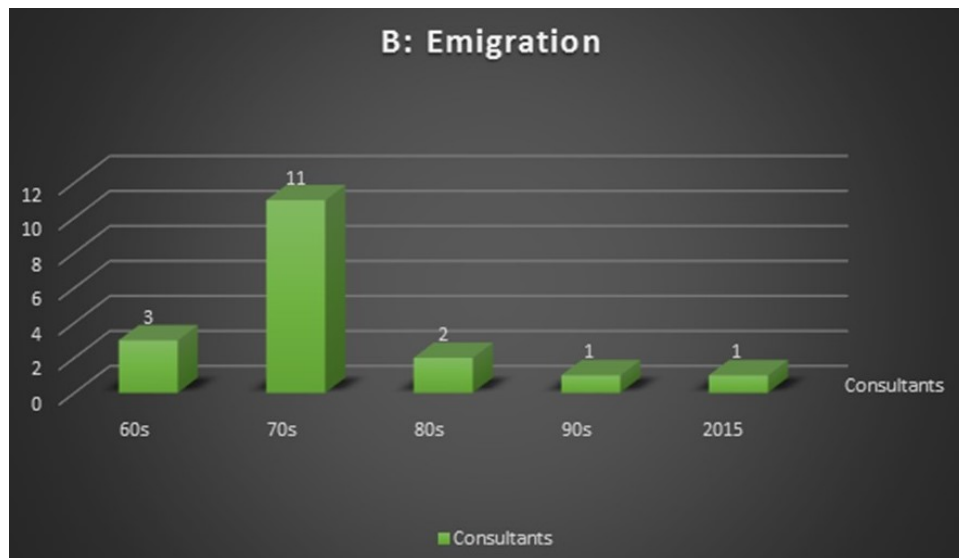


Chart B. Distribution of language consultants according to emigration period

Chart C illustrates the level of education of the consultants who were born in the village of Beth Qustan. It is clear that the majority were illiterate with no edu-

cation, whereas some found the time and had the opportunity to be schooled to a certain extent by a traditional teacher from the community in a hall belonging to the local church. Although some were given the opportunity to undertake their primary school education, very few graduated from it. The levels of education and the consequent almost nonexistent exposure to a majority language in the educational context also have a clear positive impact on the preservation of the language in the diaspora community.

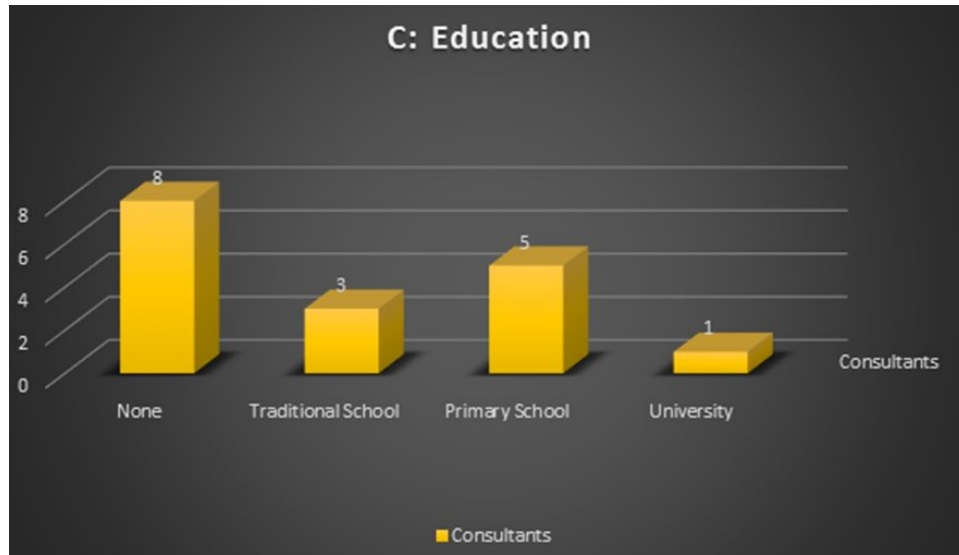


Chart C. Distribution of language consultants according to education levels

2.3 The Turoyo Language The Modern Aramaic language of the mountainous region of Tur ‘Abdin, South Eastern Turkey, is known to its native speakers as Surayt or Turoyo, that is, ‘the language of the Tur ‘Abdin’. It belongs to the Central Neo-Aramaic (CNA) language group. This group of languages is sometimes also referred to as North Western Neo-Aramaic (Nwana).

The Turoyo language is spoken by Christians living in Tur ‘Abdin and the surrounding areas. The vernacular dialects are, by definition, not written down (until modern times). It is only recently that scholars have started to look for evidence of their earlier use. Therefore, it is rather difficult to ascertain precisely how far back Turoyo was being spoken as a distinct language. Turoyo is related to North Eastern Neo-Aramaic (NENA) dialects spoken by other ancient Christian and Jewish communities of Eastern Turkey, Northern Iraq, and North Western Iran. As for intelligibility, speakers of Central Neo-Aramaic cannot easily understand speakers of North Eastern Neo-Aramaic.

The materials collected within the present project sporadically include other surrounding languages such as Syriac, German, Kurdish, Turkish, Arabic, and Persian. The literary language of the Turoyo speakers is Syriac, and it is the main ecclesiastical

language of all Aramaic dialects. Words from this language are used as it is often perceived and claimed by the natives that Turoyo is an ‘impure’ language and a corrupted form of Syriac, which has been contaminated and riddled with words borrowed from foreign languages, in particular Kurdish, Turkish, Arabic, and Persian. Attempts by natives are thus made to filter and purge Turoyo of such loan words, and to bring the spelling and use of Turoyo words into line with the “original” Syriac. However, it is common sense that this is the case in all living languages – they all contain borrowings from their neighbors. This is not a sign of weakness or corruption, but evidence of the strength, health, and vitality of a language. A language that borrows words confidently, without being threatened or weakened, is a reflection of its own ability to absorb useful additions to its vocabulary. Diaspora communities have gone a step further to include borrowings from their new “homelands”, and such is the case in this collection, for which fieldwork was undertaken in Germany.

3. Data collection The data for this project were collected in August and September 2016 in Gütersloh and Kirchardt, Germany. Notes on cataloguing metadata about the language consultants and the communicative event in general were taken immediately at the beginning of each recording session. This information was extended after the fieldwork to include descriptive, structural, technical, and administrative details, reflecting not only the conditions and contents of each recording and the analysis done afterwards with the data, but also the requirements of archiving language and cultural materials and making them available to the general public.

Ethical issues play a decisive role throughout this project.¹ In this sense, the project, its outcomes, and the meaning of having the data available online were explained to the language consultants before each recording session, and those who decided to participate in the project have signed an informed consent. The signed forms have also been archived and can be found in the relevant individual bundles.

Being a native speaker of the language, I was able to interact with the language consultants without the help of a documentation facilitator. Thus, the majority of the recordings comprise staged communicative events in the form of monologues and dialogues. No elicitation was necessary and I could focus on the collection of naturalistic data. Even though the recording setting may have lost its naturalness through the presence of the recording equipment, the recordings represent spontaneous language use in everyday communication.

In order to achieve high-quality recordings and guarantee their perpetuation over time, as well as to comply with archiving standards, the following equipment was used in the field:

1. Zoom H4n for audio collection: The recordings were done in linear PCM with the settings 48 kHz, 16 bit and stereo, and later synchronized with the video using Adobe Premiere to provide better audio for the video recordings.

¹Cf. Research Ethics at SOAS, University of London. <https://www.soas.ac.uk/researchoffice/ethics>.

2. Canon HG10 for video recordings: The camcorder produced high-quality uncompressed video files (MTS) which were later compressed to MP4 with Adobe Premiere for further ELAN processing and online streaming reasons.
3. Panasonic DMCFZ45 for JPG pictures.

Table 1 summarizes the data collected for the Beth Qustan dialect project.

Table 1. Data collected during the fieldwork in Germany

Data type	Number of files	Duration
Video	606	26 hrs
Audio	606	26 hrs
Pictures	836	n/a
Informed consent	18	n/a

4. Data processing Upon completing the fieldwork, all the video recordings representing the basic resource for each bundle in the Turoyo collection were analyzed and classified in terms of content and then edited with Adobe Premiere according to the topics discussed. A list of different topics was created, which is explained below. This initial classification was for the organization of the materials collected in the field, their naming and bundling process.

Thus, files are named with the keyword ‘Qustan’ that has remained the same throughout the project, and this is followed by a three-digit number and the file extension (e.g., Qustano63.mp4). Files derived from each other or belonging together share the same file name root, with only the extension differentiating them (e.g., Qustano63.mp4, Qustano63.wav, Qustano63.eaf, Qustano63.pfsx, Qustano63.pdf and Qustano63.jpg). Identical extensions providing a series of photographs or documents share the same first part of the file name followed by a hyphen and a number (e.g. Qustan615-1.jpg, Qustan615-2.jpg, Qustan615-3.jpg). The files were renamed using the tool File Renamer Basic.² The file name root followed by the number (e.g., Qustano63) is also the unique ID for the individual bundle.

The metadata have been created using CMDI Maker³ and Arbil (Withers 2012). Metadata for each bundle include a general description of the bundle and its content, information about the actors, demographics, location, language, genre, resources, and their access definitions. In order to facilitate a navigation within the collection, keywords and the previously defined topics were assigned to the metadata (see §5). The decision to add this information to the metadata is enforced by the fact that it allows users to undertake more structured searches, and at the same time, it fosters data analysis in the future. Figure 4 illustrates a search in the Turoyo collection according to a specific topic ‘Commodity’, which has then been further specified with the keyword ‘Electricity’ (Figure 5).

²File Renamer Basic. http://sherrodcomputers.com/products_filerenamer.cfm.

³CMDI Maker. <http://cmdi-maker.uni-koeln.de>.

Collection Turoyo

https://elar.soas.ac.uk/Collection/M1103568?types=AllFields&filter[NOT]=deletion_message&filter[resource_keyword_string]=A%20Commodity%20(AllApplied%20)

Endangered Languages Archive at SOAS University of London

Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo

Deposit | Rawfiles and resources

Showing 1 - 10 of 23 items

Before Electricity

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
Lahlo tells us about the hardships before they had electricity in the village.
Recorded on: 2016-09-05
Keywords: Turoyo - English - Narrative - Commodity - Hardship - Food - Heating - Lahlo Ajman

Clothes in the Village

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
De la tells us that they used to sew their own clothes in the village before his time. However, people began buying clothes in Hadyat during his time.
Recorded on: 2016-09-05
Keywords: Turoyo - English - Narrative - Commodity - Clothes - Hadyat - Sa Ostad

Clothing from Hadyat

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
Devin tells us that they were buying their clothing from Hadyat at his time. However, before his time they were making their clothing in the village from wool. However, he tells us that although there was growth before his time, love and respect was abundant.
Recorded on: 2016-09-05

Figure 4. Filtering data in the collection according to the topic “Commodity”

Search Results - "Electricity"

https://elar.soas.ac.uk/Search/Results?lookfor="Electricity"&type=Resource&keyword

Endangered Languages Archive at SOAS University of London

Search: "Electricity"

Showing 1 - 6 of 6 for search: "Electricity". Query time: 0.07s

Cooking Food

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
Suleyman explains how they used to cook food before electricity was provided in the village.
Recorded on: 2016-09-01
Keywords: Turoyo - English - Narrative - Traditional Foods - Cooking - Food - Electricity - Suleyman Ajman

Djins before the Introduction of Electricity

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
Suleyman explains that djins used to exist in Beth Qustan. He tells us that he used to see their fire from distance. However, they would vanish when approaching them. Nevertheless, he does not believe that they exist in Beth Qustan since electricity was introduced to the village. He explains that djins are frightened of iron.
Recorded on: 2016-09-05
Keywords: Turoyo - English - Narrative - Djins - Iron - Djins - Gevra Akut

Electricity

Deposit title: Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo
Lahlo tells us that electricity was introduced to the village during his time. However, before his time they were using oil lamps for lighting.
Recorded on: 2016-09-05
Keywords: Turoyo - English - Narrative - Electricity - Lahlo Ajman

How to use search

You can search in two ways:

- enter text in the search box and press "Search". Search is not case sensitive.
- to find variations of words add a wildcard symbol, e.g. "india*" finds "India" and "Indian".
- click a category in the left panel to filter out a set of bundles or responses.
- click on the red X next to a selected category to delete a selected keyword or facet and see all bundles or responses again.

To refine your search:

- enter two or more words for results containing all those words; e.g. entering "nigeria" and "audio" finds the deposit containing wordlists which includes recordings made in northern Nigeria.
- use the keywords in the left panel to browse and select further categories; e.g. if you search for "nigeria" and "audio", click on a language name to filter your results for recordings where this language is used.

Figure 5. Further specification of the topic search according to the keyword “Electricity”

Moreover, the recordings were segmented, transcribed, translated into English, and glossed using ELAN (Wittenburg et al. 2006) and SIL Fieldworks Language Explorer (FLEX).⁴ The annotation process, an ongoing task, was divided into three different phases. First, all the recordings have been segmented using ELAN, which allows time-aligned segmentations. In the second phase, a sample of the recordings has been orthographically transcribed (using the Turoyo alphabet described in §4.1) and translated into English, also in ELAN. Thereafter, the transcriptions were imported into FLEX and glossed morphologically.

More details on the alphabet used for the transcriptions (§4.1) and on further annotations (§4.2) will be given in the following sections.

⁴Fieldworks Language Technology. <http://software.sil.org/fieldworks>.

4.1 Turoyo Alphabet and the Transcriptions There is currently a fierce debate in some parts of the diaspora over the question of which script should be used to write Turoyo – the traditional Syriac alphabet of 22 consonants and 5 vowels, written from right to left, or a modification of the European “Latin” alphabet, written from left to right.

In my opinion, there is no need for resentment, hostility, and confrontation. The glory of Classical Syriac will never diminish. It is the language of the Scriptures and of the liturgy of the Syriac churches; it is the literary language of the Syriac intelligentsia and has been chosen and used by the greatest writers, poets, and thinkers among the Aramaic-speaking people. The status of the Syriac script is guaranteed by its continued use in the liturgy. The language will not be threatened by the adoption of the Latin script for Turoyo. The Latin alphabet is more familiar than any other alphabet for members of the Syriac communities, both in Tur ‘Abdin and in the diaspora. Therefore, the use of the Latin alphabet would greatly help their learning of the language, and will also encourage them to use Turoyo in written communication: letters, emails, SMS texts, and social media.

A modified Latin alphabet created between 2012 and 2013 and published in 2014 (Oez 2014:vii) has been used to transcribe the recordings of this documentation. Figure 6 presents the alphabet with examples.

A sample of the recordings (more than six and a half hours) has been transcribed in ELAN. Considering that transcriptions were to be further analyzed in FLEx, the ELAN files were set up in a way to easily allow transfer of data between ELAN and FLEx, using the ELAN-FLEx-ELAN workflow developed by Gaved and Salfner (2014). Thus, in the first phase, the ELAN files had a group of three main tiers – one for the transcription, one for the free English translation and one for words, all marked according to the participant being transcribed (A, B, etc.) (Figure 7).

Using the alphabet described above, a narrow orthographic transcription was created. Filled pauses and interjections were also transcribed. Incomplete words were additionally marked with |, which opens up the possibility of further research on domains like topic change strategies.

4.2 Annotations A sample of the transcriptions, corresponding to a total length of three hours, was imported into FLEx and further annotated. Apart from word-by-word English translations, part of speech information and morphological glossing were added at the morpheme level. Thus, interlinear versions of the transcriptions were created, exported from FLEx as PDFs and added to each individual bundle in the collection.

LETTER	EXAMPLE	GLOSS	AS IN	SYRIAC	ARABIC
'	'omar	to say	' (glottal stop)	ܐ	ء
‘	‘ayno	eye	‘	ܐܝܢܐ	ع
A	ar‘o	land	allowance	ܐܪܘܫܐܝܡܐ	أ
B	bayto	house	box	ܒܝܬܐ	ب
C	curho	wound	joy	ܥܘܪܗܘܐ	ج
Ç	çaye	tea	champion	ܥܝܫܐ	چ
D	darbo	road	dog	ܕܪܒܘܐ	د
Ḍ	ḍa‘ifo	slim	that	ܕܘܟܐܢܐ	ذ
Ḑ	Ḑarbo	hit, blow	those	ܕܘܟܐܢܐ	ظ
E	emo	mother	esoteric	ܐܡܐ	إ
F	femo	mouth	four	ܦܝܡܐ	ف
G	gamlo	camel	go	ܓܡܠܐ	گ
Ĝ	ĝalabe	much	gh	ܓܠܒܐ	غ
H	harke	here	here	ܗܪܟܐ	ه
Ḥ	ḥamro	wine	habibi	ܚܡܪܐ	ح
I	işmo	name	open	ܝܫܡܐ	كسرة مُخففة
İ	iḏo	hand	meat	ܝܕܐ	إي
J	jeton	token	measure	ܝܬܘܢܐ	ز
K	kalbo	dog	cat	ܟܠܒܐ	ك
Ḷ	ḵabro	word	Bach	ܟܝܠܒܐ	خ
L	lahmo	bread	little	ܠܗܡܐ	ل
M	man?	who?	man	ܡܢܐ	م
N	nuro	fire	no	ܢܘܪܐ	ن
O	ono	I	over	ܐܢܐ	او (واو مُخففة)
P	pħine	kick	people	ܦܚܝܢܐ	پ
Q	qhuto	coffee	Qatar	ܩܗܘܬܐ	ق
R	rišo	head	room	ܪܝܫܐ	ر
S	susyo	horse	sit	ܫܘܫܘܐ	س
Ş	şin‘a	profession	sore	ܫܝܢܐ	ص
Ş	şimšo	sun	sharp	ܫܝܡܫܐ	ش
T	tar‘o	door	tap	ܬܪܐ	ت
Ṭ	ṭabuto	steady	thunder	ܬܒܘܬܐ	ت
Ṫ	ṭowo	good	determined	ܬܘܘܐ	ط
U	urd‘o	frog	group	ܘܪܕܐ	أو
W	wardo	rose	wing	ܘܪܕܐ	و
Y	yawmo	day	yes	ܝܘܡܐ	ي
Z	zamoro	singer	zebra	ܙܡܪܐ	ز

Figure 6. Modified Latin alphabet for Turoyo

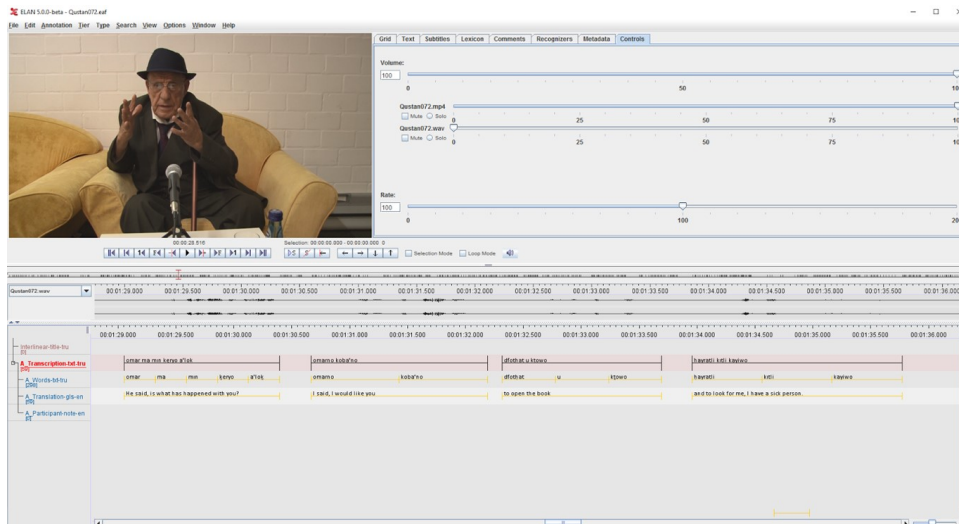


Figure 7. Transcription and translation of Qustan072 in ELAN (prior to further annotations in FLE_x)

1	Word	ono	na'imowayno			
	Morphemes	ono	na'im	-o	-way-	-no
	Lex. Entries	ono	na'imo	-o ₁	-way-	-no
	Lex. Gloss	1sg	child	sg.m	Past	1sg.m
	Lex. Gram. Info.	pro	n	n(suffix)	v:(Tense)	v:(PNG)

Free I was small.

2	Word	yani	ḡalabe	na'imowayno		
	Morphemes	yani	ḡalabe	na'im	-o	-way- -no
	Lex. Entries	yani	ḡalabe	na'imo	-o ₁	-way- -no
	Lex. Gloss	that is	much	child	sg.m	Past 1sg.m
	Lex. Gram. Info.	adv	quant	n	n(suffix)	v:(Tense) v:(PNG)

Free namely, I was very small.

3	Word	kale	ḥoli		
	Morphemes	kale	ḥol	-i	
	Lex. Entries	kale	ḥolo	-i ₁	
	Lex. Gloss	interj	uncle	1sg.n	
	Lex. Gram. Info.	interj	n	poss:(suffix lim)	

Free Here's my uncle.

Figure 8. Example showing interlinear glossing of the transcript Qustan138 (Oez 2017b)

The last step was the update of the previous ELAN files (exemplified in Figure 7) with the annotations from FLE_x using the ELAN-FLE_x-ELAN workflow. Tiers for the morphemes, their types, glossings and part of speech, as well as for segment numbering, were automatically added to the initial ELAN files (Figure 9).

Currently, two different types of ELAN files can be found in the archive, reflecting different stages of the annotation process, i.e., there are bundles with transcriptions only (approximately three and a half hours) and bundles with fully transcribed, translated, and annotated ELAN files (approximately three hours).

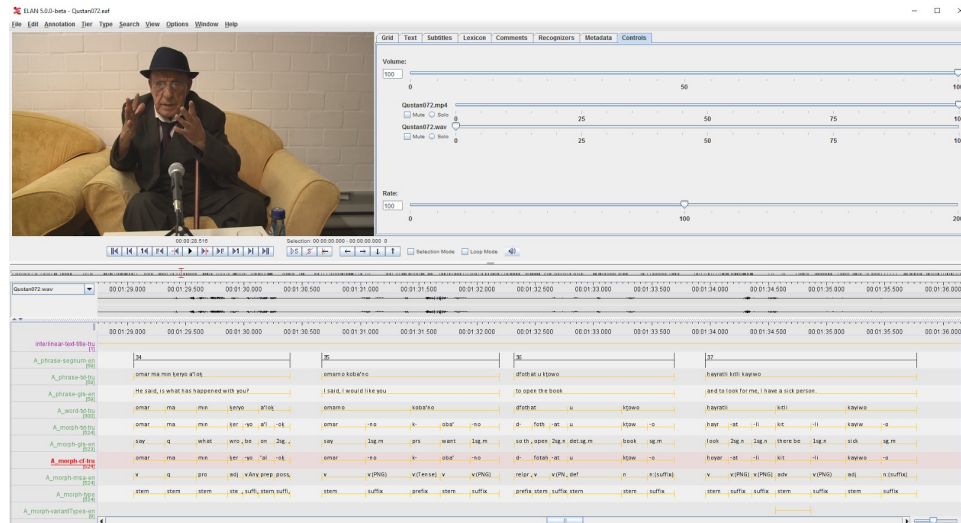


Figure 9. Full annotation of Qustano72 in ELAN after FLE_x (Oez 2017a)

5. Deposit Contents The deposit comprises 608 bundles of the Beth Qustan dialect that reflect spontaneous language use in the form of monologues and dialogues. Even though the fieldwork was undertaken with the diaspora community in Germany, the recordings have captured socio-cultural practices common for the community in the Beth Qustan village, with a wide range of themes such as legendary stories traditionally told at village gatherings, historical narratives, personal stories, explaining food production (e.g., of raisins), and the sophisticated culinary art of cooking cultural food.

All the materials are openly accessible, without any kind of restrictions, in the online catalogue of the Endangered Languages Archive at the following URL: <https://elar.soas.ac.uk/Collection/MPI1035085>. Figure 10 illustrates the deposit page of the documentation of the Beth Qustan dialect of the Central Neo-Aramaic language, Turoyo, where a first impression on the language can be gained through a podcast and a subtitled show reel. Additionally, the deposit page provides information about the project, the language, the community, and the materials collected.

Each bundle contains a video and an audio recording with picture(s), ELAN files, and PDF documents with interlinear glossing (in the case the transcriptions have

been further annotated in FLE_x) and consent forms. The total length of the videos is 26 hours. Even though not all the video recordings in the Beth Qustan deposit have been transcribed, they have all been segmented in ELAN and are available to be openly accessed by any users of ELAR, without restrictions, for transcription and/or for further research and analysis. The main reason behind the segmentation of all recordings lies in their future usability. Six and a half hours of the recordings have already been transcribed, and more than three hours have been provided with free translations in English and are fully annotated (morphologically).

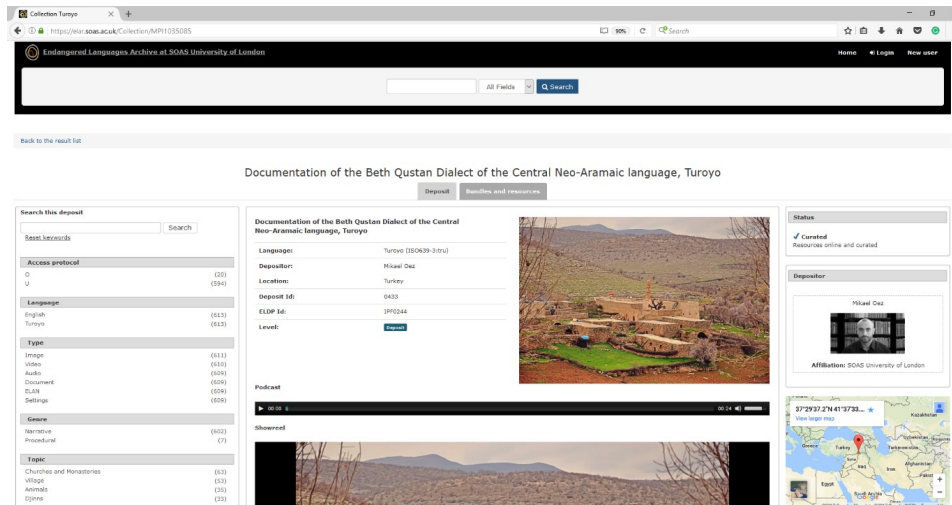


Figure 10. Deposit page of the Beth Qustan dialect collection

A search feature exists in the deposit page where users can search for materials in the collection. As mentioned above, the search can be made by filename, keywords and topics (see Figures 4 and 5 above).

As described in §4, the bundles have been categorized by topics to make the navigation more manageable when consulting specific themes. Figure 11 below indicates where the topics can be found on the deposit page.

The topics, with a number of bundles for each topic, presented in the collection are as follows: Churches and Monasteries (62), Village (53), Animals (35), Djinns (33), Marriage (26), Commodity (23), Immigration (23), Sorcery (22), Festivals (20), Story (19), Family (18), Language (18), Agriculture (17), Doctorship (15), Entertainment (15), Traditional Foods (15), Fallouts (13), Famine (12), Treasures (12), Army (11), Hardships (11), SOS (11), Childhood (10), Homeland (9), Neighboring Villages (9), Village Raid (8), Aprakhe, Kutle, Shamburake (7), Mukhtar (7), Tanoor (7), House (6), Education (5), Funeral (5), Amulets (4), Conversion (4), Political Parties (4), Livelihood (3), Construction (2), Yazidis (2).

The topics have been intentionally chosen to focus on sociolinguistic aspects of the Beth Qustan village. I am going to illustrate those aspects and the potentials of topic classification for data analysis through the recordings included in the topic “Aprakhe, Kutle, Shamburake”. Nazan Akkaya, Hana Demirel and Hazniye Agirman

(see Figure 12) were enlisted to represent a scenario of how women would get together to cook dinner. Qusnean women usually opt out of being recorded, as traditionally only men speak at village gatherings. These women agreed to be recorded cooking while telling us, spontaneously, about life in the village, with a bit of side gossip. It is customary to talk about many things when women get together and cook traditional dishes.

Topics

- Churches and Monasteries (53)
- Village (53)
- Heritage (53)
- Clubs (53)
- Marriage (26)
- Comedy (23)
- Immigration (23)
- Barney (23)
- Festivals (20)
- Story (18)
- Family (18)
- Language (18)
- Agriculture (17)
- Entertainment (16)
- Contesting reflections (15)
- Dialectship (15)
- Traditional foods (15)
- Introduction (14)
- Religion (13)
- Parade (13)
- Treasures (12)
- Army (11)
- Hardships (11)
- SOB (11)
- Childhood (10)
- Heritage (9)
- Neighbouring villages (9)
- Village Road (9)
- Alphabetic, Kufic, Shamsanic (7)
- Musical (7)
- Teaser (7)
- House (6)
- Education (6)
- Funeral (6)
- Political parties (6)
- Animals (6)
- Conversion (6)
- Localhood (6)
- Construction (6)
- Trade (6)
- Use (6)

Participants

- Dulaman Agirman (114)
- Eliza Akar (104)
- Iskender Demirel (93)
- Lilwa Agirman (92)
- Lebla Isakban (77)

Summary of deposit

A dialect of Neo-Aramaic spoken in the Christian village of Beth Qustan or Beth Qustan since time immemorial. It is known in Neo-Aramaic as 'the Qusnean', in Syriac: Beth Qustan (literally 'the house of Constantine', in Kufic: 'Baqstan', and it has become known by the Turkish name 'Rigat' since the 1920s. Beth Qustan is one of dozens of villages in Tur Abdin (literally in Syriac: 'the mountain of worshippers') speaking a Neo-Aramaic dialect. It is one of the few remaining Christian villages within the Christian stronghold of Tur Abdin, and it is a mountainous sector of the Hatay Province, South Eastern Turkey. Beth Qustan was a flourishing rural centre and part of the food basket of the Fertile Crescent at the turn of the 20th century, however, only an estimated 20 families remain inhabitants of Beth Qustan in 2017. The Qusnean still speak a specific dialect of the Central Neo-Aramaic (CNA), which is better known to the community as 'Turoyo', the language of Tur Abdin.

Beth Qustan, like many other villages in the Tur Abdin region, was probably inhabited in pre-Christian times, and it is believed that the village was a place of rest for travellers and the army of Constantine I regularly marching through the village. Located near the village is the rock of Hajar which was named after St Hajar, the mother of the Syrian Emperor, Constantine I. Beth Qustan has nursed the renowned saint St Gabriel, who was born in 374 who dedicated his life to the Monastery of Qastamin. Following the performance of several miracles, St Gabriel became the patron saint of the monastery of Qastamin, which since the end of the 12th century has become known as the St Gabriel Monastery.

The project comprises socio-cultural practices of the Turoyo speaking community in Tur Abdin, focusing on vernacular tales, particularly those that demonstrate cultural interaction between Muslims and Christians, including Muslim visitors to the shrines of Christian saints, and consultation of soothsayers by Christians.

Further information on the documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language can be accessed on the ELAK blog comprising an article authored by Hicret Oez following the Feldwork undertakes in Gülebişik and Kiribasi in Germany.

Group represented

The group represents the community of Beth Qustan, Neo-Aramaic dialect of the Beth Qustan (House of the Holy Father) in Hatay Province, Turkey, who have been documented and collected through the project.

Figure 11. Topics in the Beth Qustan dialect collection



Figure 12. (From left to right) Nazan Akkaya, Hazniye Agirman, and Hana Demirel

They decided to cook three dishes, namely *aprakhe*, *kutle*, and *shamburake*. They begin with explaining how the ingredients have been prepared, and proceed with explaining how they are made and cooked.

Nazan introduces herself and explains how she has prepared the ingredients to cook *kutle* (see Figure 13), which is a dish made of wheat balls stuffed with minced meat and chopped vegetables. Hazniye also introduces herself and explains how she has prepared the ingredients to cook *shamburake* (see Figure 14), a dish made of folded bread dough with minced meat and chopped vegetables. Hana (on the right) also introduces herself and then explains how she has prepared the ingredients to cook *aprakhe* (see Figure 15), which is a dish made with minced meat, rice, and chopped vegetables wrapped in grape leaves.



Figure 13. Kutle



Figure 14. Shamburake



Figure 15. Aprakhe

In fact, the recordings described above represent standard procedural events in the domain of traditional cooking. However, they exemplify not only traditional recipes and their preparation, but also reflect the social structure behind this kind of communitarian event. Moreover, their potential goes far beyond the mere documentation, considering that their value is unquestionable in terms of language maintenance and possible revitalization activities.

6. Consulting the Data and Reference Anyone can deposit digital documentation of endangered languages with ELAR, at SOAS, University of London.⁵ A licensing agreement, listing copyright rules, is completed and signed by the depositor to permit ELAR to disseminate the deposit for remote, world-wide research and education under the terms of a Creative Commons Attribution/Non-Commercial/NoDerivatives (CC BY-NC-ND 4.0) international license. All uses shall be consistent with the terms of U.K. copyright legislation's fair use (also known as "fair dealings" and "fair practice") provisions.⁶

Users of any part of the collection should acknowledge Mikael Oez as the principal investigator, the data collector and the researcher. Users should also acknowledge the Endangered Languages Documentation Programme, ELDP, as the funding body of the project. Individual speakers whose words and/or images are used should be acknowledged by their respective name(s). Any other contributor who has collected, transcribed or translated the data or was involved in any other way should be acknowledged by name. All information on contributors is available in the metadata.

To refer to any data from the corpus, please cite as follows:

Oez, Mikael. 2017. Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo. ID: Qustan[insert ID number here]. London: SOAS, Endangered Languages Archive, ELAR. (<https://elar.soas.ac.uk/Collection/MPI1035085>) (Accessed [insert date here].)

7. Project Outcomes and Further Research The project is designed to shed some light on establishing more precise boundaries between the dialects and on the dynamics of language contact. This will introduce much-needed fresh material to boost discussion about the mechanisms of interaction between languages, such as lexical borrowing and externally induced grammar. For instance, villages with better links and nearer to the city of Midyat, the main urban center, are often more influenced by the Arabic language, and Arabic words are often Aramaicized in Turoyo, i.e., they take an Aramaic pattern when they are conjugated, whereas villages borrow from their neighboring Kurds, and hence Aramaicize Kurdish words. The Aramaicizing of loanwords is particularly interesting when they are verbs, as they match the endings of the native words when they are conjugated.

All project outcomes have been made available in the online catalogue of ELAR. They consist of a multimedia time-aligned, transcribed, translated, and annotated corpus of Turoyo; a dictionary based on the transcriptions and glossing of 66 recordings, with a total length of three hours of recordings, corresponding to 4,612 types and 19,794 tokens; a sketch grammar based on 66 recordings, with a total length of three hours of recordings; a list of idioms and proverbs extracted from the corpus; and multiple articles reflecting on different analysis of the collected data.

⁵ELAR, SOAS, University of London. <https://www.soas.ac.uk/elar/depositing-with-elar>.

⁶UK©CS. The Copyright Service. http://www.copyrightservice.co.uk/copyright/p27_work_of_others#fair_dealing.

8. Concluding Reflections Central Neo-Aramaic dialects form a very diverse group of Aramaic dialects that were spoken until modern times in South Eastern Turkey by Christian communities. These are among the last remaining living residues of the Aramaic language, which was one of the major languages of the region in antiquity. There are currently at least a couple of dozen Neo-Aramaic dialects which originated from Tur 'Abdin, that have not yet been documented.

A construction of a FLEx database is imperative to produce grammar sketches of the individual dialects and make these sketches available to the public. It will make it possible to do comparative searches across the dialects for analytic purposes, and to create customized electronic map displays of the distribution of linguistic features.

Because many speakers of Central Neo-Aramaic dialects have been forced to emigrate, most of the dialects are now in danger of extinction. It is, therefore, an urgent task to study and document the dialects while competent speakers can still be located, and before they are extinct. I realized during my fieldwork that if these are not documented prior to losing the generation of native speakers I worked with, that is, within five to 10 years, we will also lose all invaluable knowledge about their culture and traditions. Aramaic civilization goes back more than 3,000 years, which we have regrettably not documented. We have the technology and equipment today to document this fascinating civilization at a very low cost. We do not, however, have time on our side.

References

- Bouda, Peter, Vera Ferreira & António Lopes. 2012. Poio API – An annotation framework to bridge language documentation and natural language processing. In Mambri, Francesco, Marco Passarotti & Caroline Sporleder (eds.), *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities*, 15–26. Lisbon: Colibri.
- CMDI Maker. <http://cmdi-maker.uni-koeln.de>. Accessed 05-06-2017.
- The Endangered Languages Archive (ELAR), SOAS, University of London. *Depositing with ELAR*. <https://www.soas.ac.uk/elar/depositing-with-elar>. Accessed 05-06-2017.
- Gaved, Tim & Sophie Salfner. 2014. Working with ELAN and FLEx together: An ELAN-FLEx-ELAN teaching set. <https://tla.mpi.nl/tools/tla-tools/elan/thirdparty>. Accessed 06-06-2017.
- Himmelman, Nikolaus. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus Himmelman & Ulrike Mosel (eds.), *Essentials of language documentation*, 1–30. Berlin: Mouton de Gruyter.
- Oez, Mikael. 2017a. Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo. ID: Qustano72. London: SOAS, Endangered Languages Archive, ELAR. <https://elar.soas.ac.uk/Record/MPI1044257>. Accessed 07-06-2017.

- Oez, Mikael. 2017b. Documentation of the Beth Qustan Dialect of the Central Neo-Aramaic language, Turoyo. ID: Qustan138. London: SOAS, Endangered Languages Archive, ELAR. <https://elar.soas.ac.uk/Record/MPI1044857>. Accessed 07-06-2017.
- Oez, Mikael. 2014. *Modern Aramaic in practice: Turoyo: The language of Tur 'Abdin*. London: Modern Aramaic Press.
- Sherrod Computers. *File Renamer Basic*. http://sherrodcomputers.com/products_file-renamer.cfm. Accessed 05-06-2017.
- SIL International. *Fieldworks Language Technology*. <http://software.sil.org/fieldworks>. Accessed 06-06-2017.
- SOAS, University of London. *Research Ethics at SOAS*. <https://www.soas.ac.uk/researchoffice/ethics>. Accessed 05-06-2017.
- UK©CS. The Copyright Service. http://www.copyrightservice.co.uk/copyright/p27_work_of_others#fair_dealing. Accessed 05-06-2017.
- Withers, P. 2012. Metadata management with Arbil. In Arranz, V., D. Broeder, B. Gaiffe, M. Gavrilidou & M. Monachini (eds.), *Proceedings of the workshop Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LR at LREC 2012*, 72–75. Nijmegen, The Netherlands: European Language Resources Association (ELRA). Max Planck Institute for Psycholinguistics: The Language Archive. <https://tla.mpi.nl/tools/tla-tools/arbil>. Accessed 05-06-2017.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 1556–1559. <https://tla.mpi.nl/tools/tla-tools/elan> Accessed 06-06-2017.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. In Austin, P. (ed.), *Language documentation and description*, vol. 1, 35–51. London: SOAS.
- Woodbury, Anthony C. 2014. Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In Nathan, David & Peter K. Austin (eds.), *Language documentation and description*, vol. 12, 19–36. London: SOAS.

Mikael Oez
mo14@soas.ac.uk