

Endangered domains, thematic documentation and grammaticography

Jacob Terrell

University of Hawai'i at Mānoa

When setting out to document a language with the intended goal of describing it (typically through a grammar and dictionary), fieldworkers prefer to collect an array of linguistic data, ranging from elicited words and paradigms to an assortment of texts based on conversations, narratives, procedures and so forth. Capturing a wide variety of speech acts provides a clearer record of the language and its use, and thus offers the potential for a richer description of the language at hand. However, without controlling for content, one may collect linguistic data based on an open-ended amount of topics or themes. The purpose of this chapter is to introduce the notion of endangered linguistic domains and themes in language documentation and description. Even in thriving minority languages, domains such as indigenous music or knowledge of flora and fauna come under pressure from the same forces that eventually lead to language endangerment. Gathering linguistic data based on a particular domain or specialized knowledge can generate a corpus applicable to a wider audience without sacrificing the needs of linguists. Similar to thematic dictionaries in lexicography, this introduces thematic grammars to grammaticography.

1. INTRODUCTION.¹ The purpose of this chapter is to introduce the notion of endangered linguistic domains as they pertain to thematic documentation and grammaticography through corpus building. When setting out to document a language with the intended goal of describing it (typically through a grammar and dictionary), fieldworkers often collect an assortment of texts of various genres (Himmelman 2006). Capturing a wide variety of speech acts in a corpus provides a clearer record of the language and its use, and thus offers the potential for a richer description of the language. For example, Mosel (2011b) provides a useful overview of what fieldwork guides recommend one collects in terms of data, including recording different genres such as oral histories, narratives, explanatory texts, artistic texts (songs, poems, etc.) and so on. However, without controlling for domain or topic, one may collect linguistic data based on an open-ended number of subjects.

In this chapter I argue that in some cases it is appropriate to control for topic and focus on a single domain or genre in a documentation project. This is a novel concept, and so it leads us to ask how one should go about thematic documentation of an endangered domain, how this approach may have an impact on a grammar, and how the outputs of such a project can be of use to the community, linguists, and others. To address this, Section 2 raises the notion of endangered linguistic domains. The following section introduces the idea of thematic documentation for corpus building and grammaticography. Following this, I will present my work with the Akha shaman of northern Thailand as a case study, discussing

1 I would like to thank the editors of this volume, Toshihide Nakayama and Keren Rice, as well as an anonymous reviewer for their comments on previous versions of this chapter. I am also grateful to Pirma Gavq Lavq for sharing the Ahka shaman chants with me, and to Miqder Saeduq for her constant assistance. Last, the case study described below would not have been possible without support from the Hans Rausing Endangered Languages Project and the World Oral Literature Project.

how the thematic documentation project has shaped the resulting corpus, dictionary, and grammar.

2. ENDANGERED DOMAINS. Even in thriving minority languages, domains such as indigenous music or knowledge of flora and fauna come under pressure from the same forces that eventually lead to overall language endangerment. In fact, one may presume that language endangerment is a systematic process where individual domains first become moribund and then disappear, gradually leading to the endangerment of the language as a whole. This is reflected in the fact that before becoming moribund, the domains of endangered languages are reduced from every type of social application down to the home domain. At the time of writing, there is no definition of an endangered domain or genre, nor any rubric to measure the level of endangerment as is often applied to languages as a whole. However, the concept is beginning to receive some attention.

For example, in a post titled *Endangered Genres*² on the *Endangered Languages and Cultures* blog hosted by Paradisec, Peter Austin writes:

It is by now well known that around half (or possibly more) of the world's 7,000 languages are endangered and under threat of disappearance during the current century. Perhaps less well known is that many languages that are not (yet) endangered show certain genres, or ways of using the language, that are endangered in that there are few people who can perform them and occasions for their use are diminishing. We could refer to these as 'endangered genres'.

Austin goes on to discuss a literary tradition of the Sasak of Indonesia, where they once recorded manuscripts on lontars, the dried leaves of a type of palm, to be read during performances. These manuscripts are unique in that they are "written in Kawi, a form of middle Javanese, or Sasak, or a mixture of both". Austin further states that this literary practice, or domain, is highly endangered, "as there are probably only 100 people (among a population of 2.5 million) who can read the manuscripts, and performances are discouraged due to cultural associations which conservative Islamic groups on Lombok do not approve of".

Relatedly, Tim Brookes of the *Endangered Alphabets*³ project writes:

Writing has become so dominated by a small number of global cultures that those 6,000-7,000 languages are written in fewer than 100 alphabets. Moreover, at least a third of the world's remaining alphabets are endangered—no longer taught in schools, no longer used for commerce or government, understood only by a few elders, restricted to a few monasteries or used only in ceremonial documents, magic spells, or secret love letters.

Like the previous example of Lontar manuscripts, this is a complex case of endangered domains in that not only does it involve a single domain, that of literacy, but in some cases other domains may be endangered as well, such as chanting, magic, and certain literary practices.

2 <http://www.paradisec.org.au/blog/2010/12/endangered-genres>

3 www.endangeredalphabets.com

Finally, perhaps no organization has given more attention to the idea of endangered domains than the World Oral Literature Project (WOLP) through the University of Cambridge Museum of Archaeology and Anthropology. The project was initiated in 2009 and its goal is to record, document and archive collections of endangered oral traditions. The WOLP defines oral traditions on their website as follows:⁴

Most simply, oral literature refers to any form of verbal art which is transmitted orally or delivered by word of mouth. These creative works are increasingly endangered as globalisation and rapid socio-economic change exert complex pressures on smaller communities, often eroding expressive diversity and transforming culture through assimilation to more dominant ways of life. As vehicles for the transmission of unique cultural knowledge, local languages encode oral traditions that become threatened when elders die and livelihoods are disrupted.

All projects funded by the WOLP place an emphasis on the domain of oral literature. These include collections of chants, various literary genres, indigenous religion, epics, ballads, and others. Although still in its infancy, the WOLP is by far at the forefront of thematic documentation. Currently, the organization has funded fifteen projects that are in various stages of completion.

The Lontar manuscripts, the Endangered Alphabets project, and the projects funded by the WOLP all have two things in common. First, the projects control for topic by focusing on a particular domain; they show that language documentation is cultural documentation, a concept that many fieldworkers would agree with. More importantly, they illustrate that the relationship between language documentation and conservation and cultural documentation and conservation is not unilateral. It is common for proponents of language documentation and conservation to argue that when a language is lost, the culture is lost as well. However, the examples above demonstrate that it is equally true that when a particular cultural practice is lost (or replaced), the language associated with that practice is lost too. Thus, there is a bilateral rather than unilateral relationship between the two—a concept not often recognized. For this reason, it is important for fieldworkers who set out to document and describe a language to begin considering thematic documentation of endangered domains.

3. THEMATIC DOCUMENTATION. It is often the case that the younger generations are less likely to be well-versed in detailed cultural practices and indigenous knowledge due to the ease of replacing these with more dominant local or global ones. For example, a community may be transitioning from an indigenous religion to one of the world's major religions, or learning life skills in schools rather than in the forest, or purchasing clothing from a large department store rather than weaving one's own.⁵ Whatever the case, it is this detailed cultural or indigenous scientific knowledge that is most likely to be cherished by

4 www.oralliterature.org

5 I should clarify here that I am not suggesting that any group not have access to national education, services, or goods. Instead, once these practices disappear, so does the language and knowledge which is used in these domains. Hence, when the domain is lost without any documentation, so is the language associated with it.

elders and valued by researchers in other fields. Therefore, gathering linguistic data based on a particular domain may generate a corpus applicable to a wider audience.

Of course, a researcher setting out on a thematic documentation should choose a domain or genre related to personal interests. Although potential endangered domains for a thematic documentation project depend on the environment of the speakers, the increasingly globalized world is (unfortunately) providing more topics to choose. As such, these are nearly endless: cyclical and life-cycle ceremonies, knowledge of flora and fauna, indigenous music, religion, textiles, and so on. These are indeed hypothetical examples, but still plausible. In fact, it is not necessary for the domain in a thematic documentation project to be an endangered one. The point of thematic documentation is to control for the topic to create a database of specified knowledge, which in turn should have a wider-reaching audience than just linguists.

With that said, thematic documentation itself is not a new concept—field linguists and anthropologists have long been interested in lexical semantics and domains such as kinship terms or those related to ethnobotany, ethnobiology or folk taxonomy (Haviland 2006). The Dictionary Development Process⁶ initiated by Ron Moe at SIL is a good example of thematic documentation in that it provides a template for a lexicography project based on semantic domains. Additionally, theme-based lexicography projects also allow for quick production of thematic dictionaries, each based on a particular domain, creating a focused product that is appealing to not only the community but also researchers in other fields (Mosel 2011a).

As mentioned in section 1, typical products of a documentation project are often a grammar and dictionary, and, similar to thematic dictionaries in lexicography mentioned above, thematic documentation introduces thematic grammars to grammaticography. Of course, this means that examples in a grammar based on a theme-based corpus would be restricted to its domain. However, this does not mean that such a corpus cannot fulfill the needs of linguists. For example, if one's hobby is fishing, it would be possible to gather needed linguistic data for a grammar and dictionary based on the domain of fishing. This could include names of fish, procedural texts on making nets/traps, fishing methods, folklore involving fish, narratives about fishing, and others. A corpus of this type could be of interest to biologists, ichthyologists, anthropologists, and other researchers, depending on what the corpus contains and what is made available in the translations of the texts. Additionally, the researcher would have an excuse to spend much time in the field fishing.

4. THE AKHA SHAMAN, A CASE STUDY. To exemplify the points made above, I draw on my experiences working with the Akha shaman, or Pirma, in Northern Thailand to document the chants of the indigenous religion. Akha is a Tibeto-Burman language, belonging to the Lolo-Burmese branch (Bradley 1979). The language is spoken in five countries: southern China (est. pop. 200,000) Laos (66,100), Myanmar (200,000), Thailand

6 <http://www.sil.org/computing/ddp/>

(56,600), and Vietnam (1,260) (Ethnologue 2009).⁷ With a population well over 500,000, this speech community is much larger than those which are usually the focus of a language documentation project. The language is vibrant and is still being learned by children. However, due to recent cultural changes, the register used in the traditional religion is highly endangered; in fact, it has become moribund in the last two to three generations. According to the Mekong Akha Peace and Sustainability network, an NGO made up of community leaders who work to establish cultural networks across borders, there are only three shaman in China, fewer than fifteen in Thailand, and perhaps four in Myanmar. The number of shaman in Laos and Vietnam is unknown (Wang, pc.).

One of the most important roles of the shaman is to carry out funeral rites. A funeral can be extremely extravagant; it can last up to a week and a number of animals are sacrificed to send off with the deceased, including up to three water buffalos. During the rite, the shaman will sit in front of the coffin and chant for the deceased to take the soul to the border of the afterworld. The chants are epic poems covering topics from creation stories, to conception, death, the relationships between spirits and humans, crops, animals, and much more. There is one volume of chants for each of the three buffalos that are sacrificed at the funeral, and it can take up to a week for a shaman to perform a three-buffalo funeral. The chants are all done by memory, and all the shaman that I interviewed report that it takes up to twenty years to memorize all three volumes. Still, as mentioned earlier, this practice is highly endangered. The shaman are all elders, and I do not know of any apprentice under the age of forty.

It is important to note here that the religious register used by the shaman is not mutually intelligible with contemporary, spoken Akha. The community believes that the religious register is a fossilized version of contemporary Akha. I have played recordings of chants for numerous Akha who will recognize that it is Akha, and that it is the shaman's language, but they always report that they cannot understand it. Given the dire situation of the shaman's language, some community members suggested that I direct my attention towards documenting the religion. Thus, in the fall of 2009 I initiated a thematic documentation project of shaman's chants and indigenous religion based out of Chiang Rai, Thailand.

4.1 DOCUMENTING THE CHANTS.⁸ With the help of funding from the Endangered Language Documentation Programme (ELDP) and the World Oral Literature Project (WOLP), I established a team of five native speakers of Akha to assist with the project. Three team members were in their twenties, while two were in their fifties. The team members were trained as native-speaking documenters in basic linguistic description, the applicable tools and technology, and methods of ethnography. For linguistic description, much of the curriculum was based on what is used at the Language Documentation

7 These figures are from Ethnologue, based on Bradley 1997. However, they may not be accurate due to many Akha migrating from Myanmar and into Thailand who do not receive citizenship and are not included in any census. For example, the Akha Foundation NGO in Chiang Rai believes that there are around 100,000 Akha in the 284 villages in the northern Thailand (Kukeuwsakem, pc).

8 In the 1980s, Leo Alting von Geusau and Inga-Lill Hansson worked together to initiate a project to translate cultural texts of Akha. They began with *Oer Zar*, 'Life Cycles', as recited by Pirma Arso Dzoeqbaw and Argaw Dzoeq, and produced a manuscript in 2002. Unfortunately, the project came to a halt after Dr. Geusau's untimely death in the same year.

Training Center (LDTC) run by the Linguistic Society of Hawai'i.⁹ The team members were also taught how to use Audacity, ELAN, Toolbox for glossing texts, and Lexique Pro for building the dictionary database. We also practiced using the equipment: shotgun and lavalier microphones, Zoom recorders, and the video recorder. In all, the training session was quite intensive and lasted just over a month.

After the training session, we would take turns traveling to different villages in northern Thailand to record chants and interviews, check collected texts, and take photographs of relevant items related to the chants and the religion. We worked primarily with one shaman, Pirma Gavq Lavq of Huay Pra Sot village. Much of our time was spent annotating, glossing and translating the chants, since we were working with one source language (the religious register), and two target languages (contemporary Akha and English). We found that it would take nearly a month for two team members to process a thirty-minute text, creating a solid first draft of the text—glossed with free translations in Akha and English. After we finished one text, we would travel to two villages to check our work with at least two other shaman. We continued to collect and process data in this manner throughout the duration of the project. In all, we collected and processed the data for almost a year and a half, from August 2009 through December 2010.

5. RESULTING CORPUS.¹⁰ The data in the corpus can be divided into two categories: 1) primary audio recordings of the chants performed by Pirma Gavq Lavq, and 2) secondary audio and video recordings and texts related to the chants.¹¹ Since the purpose of this project was to record, document, and describe the songs of the Akha shaman, we focused on processing Pirma Gavq Lavq's chants at the expense of processing the secondary data.

The corpus contains 10 hours and 9 minutes of Pirma Gavq Lavq chanting *Aqnyoq tiq mawr*, 'Buffalo one', the first volume of chants performed at a funeral. In total, the chants contain over 6,500 verses of prose.¹² Each verse has been transcribed in ELAN using the standard Akha orthography. We have a free translation from the religious register to contemporary Akha for the majority of the verses (perhaps ninety-five percent), though there are some sections that neither various shaman nor we could decipher. We also have a free translation in English for about seventy percent the verses that could be translated into contemporary Akha. Every verse with an Akha and English translation is also accompanied by a gloss in both contemporary Akha and English.

6. OUTPUTS. In another chapter in this volume, Mosel discusses the importance of the types of texts one collects and methods in building a corpus for writing a grammar. Additionally,

9 <http://www.ling.hawaii.edu/~uhdoc/>

10 The corpus is currently being prepared for archiving with the Endangered Languages Archive (ELAR) at the School of Oriental and African Studies (SOAS) at the University of London, and with the WOLP. The first three chants of 'Buffalo One' are more public than the others are, and thus may be accessed in the archives without restrictions. The remaining chants may be accessed with permission from Mekhong Akha Peace and Stability (MAPS).

11 The secondary data include chants from other shaman, mourning songs performed at a funeral, interviews with elders on how to carry out funeral rites, written texts involving folklore, and other materials.

12 For a comparison, the English version of the Christian New Testament contains roughly 7,500 verses.

Rehg (this volume) argues for the need to include speakers of the target language in the documentation and description processes, providing speakers with a larger role in both. In this case study we recorded different types of texts related to religion (procedural, narratives, etc.), but translating the chants remained our priority. Furthermore, the Akha assistants were more involved with the documentation of the chants and compiling the dictionary (and are listed as coauthors), while they mainly served as consultants for the grammar.

6.1 DESCRIPTIVE OUTPUTS: DICTIONARY AND GRAMMAR. The most recent draft of the dictionary contains approximately 3,300 entries. It is a trilingual dictionary: Shaman's Akha—Contemporary Akha—English. Lexical items include everyday contemporary Akha words, their clipped forms used in the shaman song, and lexical items that are unique to the shaman's chants. Since this is a thematic documentation project, the majority of the items in the dictionary revolve around the material culture and flora and fauna mentioned in the chants. To some extent this product may be of use to the community, since it is a trilingual dictionary.

I do not believe that an academic grammar will be of great use to the community, other than perhaps adding some prestige to the language and the shaman's role. (This is because, in part, it is written in English.) Instead, the intended audience for the grammar are linguists and others scholars. The grammar is based on the first three chants of 'Buffalo One': *Lavq Khoer Mr*, 'Inner Ceremony' (ch-gl-A.txt), *Gee Jm*, 'the Ancestor Alter' (ch-gl-B.txt), and *Khoer Naevq Dav*, 'Inner Spirits' (ch-gl-C.txt). In total, these texts amount to fifty-four minutes of chanting, some 446 verses. Below are two example verses.

(1) *Khoer Naevq Dav*¹³, 'Inner Spirits' (ch-gl-C.004.txt)¹⁴

CH:	<i>Nya</i>	<i>'law</i>	<i>aq.poeq</i>	<i>ghaq</i>	<i>'or</i>	<i>zer.'zaq</i>	<i>zer.'sanr</i>
AG:	<i>Lavq.tav</i>	<i>law</i>	<i>aq.poeq</i>	<i>ghaq</i>	<i>or</i>	<i>zer.zaq</i>	<i>zer.sanr</i>
EG:	Above	PTC	ancestor	CLF.person	PTC	guardian.of.children	

CH:	<i>tiq</i>	<i>'eeq</i>	<i>'or!</i>
AG:	<i>tiq</i>	<i>eeq</i>	<i>or</i>
EG:	one	gather	PTC

Akha: *Lavqtav paw eq jawr aqpoeq nar a zerzaq zersanr tiq deq mae!*

Eng: Oh!, Above, the ancestors, the guardian spirits of the children gather together!

(2) *Lavq Khoer Mr*, 'Inner Ceremony' (ch-gl-A019.txt)

CH:	<i>sar</i>	<i>'nan</i>	<i>manq.'paq</i>	<i>dziq 'awr kawr</i>	<i>'lmr</i>	<i>tmr</i>	<i>'e</i>
AG:	<i>yaw sar</i>	<i>ar.nan</i>	<i>manq.paq</i>	<i>dziq awr yaw kawr yaw lmr</i>	<i>ngaeq</i>	<i>e</i>	
EG:	easy	day	stallion	ride and quick	warm	speak	GEN

13 Akha is a tonal language, and tones are represented in the orthography as: \$-r, high; \$-0, mid; \$-q, low. Also, \$-v indicates creaky voice.

14 CH stands for chant, or what the shaman sings in each verse, AG is the gloss in contemporary Akha, and EG is the English gloss.

CH: *ar.'nan* 'leir.
 AG: *ar.nan* leir
 EG: day NPST

Akha: *Yaw sar arnan anr manqpaq dziq awr yaw kawr aer ngaeq e arnan lei.*
 Eng: A relaxing day, it is a day for riding stallions and making quick, warm conversation.

Both registers use SOV order with modifiers following the head of the noun phrase: i.e., stative verb, degree word, determiner, and then classifier. The language has postpositions rather than prepositions. There is no difference in word order in complex constructions such as passives, causatives or relative clauses. However, the structure of the verses differs from that of contemporary Akha in the interaction between prosody, morphology, and semantics.

For example, in the data above one can see that each verse contains an odd number of syllables. Verses can range in length from five syllables to twenty-one or more. The metric foot in the chants is iambic consisting of two syllables, and the final, odd syllable of a verse is footed. Obviously, not all clauses or sentences in contemporary Akha consist of an odd number of syllables. To maintain the iambic rhythm with an odd number of syllables in the chants insertion and/or clipping is used.

The morphemes glossed as PTC in (1) above are non-lexical vocables¹⁵, or syllables in song that do not have meaning but do have function. Here, their function is to ensure that the verse contains an odd number of syllables. Although the poetic syllables above fall on the stressed beat of the foot, they may also be inserted in unstressed positions. These syllables are always [a], [o], or [u], with either high or low tones.

While insertion and clipping work in tandem to ensure that a verse has an odd number of syllables, clipping is more complex. For example, there is a strong tendency in Akha for verbs, stative verbs and closed word classes to be monosyllabic. Nouns can be disyllabic, many of which have only two syllables. Note in (2) that there are two forms for 'day' in the same verse: *nan* and *arnan*; the latter is the term used in contemporary, spoken Akha (speakers do not recognize *nan* as 'day'). Also, the clipped syllable in 'day' is always *ar-* and never *-nan*. The syllable that can be clipped in a disyllabic noun is specified for each lexical item, as shown below.

TABLE 1. FULL AND CLIPPED NOUNS IN THE CHANTS

A			B		
English	Full Form	Clipped Form	English	Full Form	Clipped Form
water	ir.cuvq	ir	gourd	ir.puq	puq
head	uq.duq	uq	horn	uq.coer	coer
banana	nga.baev	nga	bean, pod	ar.baev	baev
mouth	maeq.boeq	maeq	tongue	maeq.lar	lar
earth	mir.tsaq	mir	country	mir.khanq	khanq

15 This is the term commonly used in music. I prefer the phrase 'poetic syllable' and the adhoc gloss PTC for simplicity.

In column A above the second syllable may be clipped in the chants, while in column B the first syllable of the noun may be clipped. The option to clip either the first or the second syllable depends on what I call the semantic weight of the syllable. For ‘water’ the syllable with the semantic weight is *ir-* (retained in clipped form), while for ‘gourd’ it is *-puq*. This is a result of a noun class system in Akha where nouns in a particular semantic domain share a common syllable through extension. Other nouns in the class of water/liquids include *irbanq*, ‘cup’, *irxoi* ‘urine’, *irxmr* ‘otter’, *irnei* ‘mud’. Nouns associated with the head often begin with *uq-*, as in *uqduq* ‘head’, *uqcoer* ‘horn’, *uqghmq* ‘pillow’, and *uqlanr* ‘turban’. Similar examples can be found for the other nouns above.¹⁶ These types of extensions are frequent in the lexicon, and speakers do not recognize the clipped forms above as lexical entries.

The shaman may use either *ircuvq* or *ir* for ‘water’, or *irpuq* or *puq* for ‘gourd’, as long as the syllable with the semantic weight falls on the stressed beat in the iambic foot, *ir* and *puq* respectively. If the syllable with semantic weight is not aligned with the stressed beat of the foot, clipping, insertion or a combination of the two can be used.

Returning to ‘day’ in (2), the syllable with the semantic weight, *nan*, falls on the stressed beat of the foot; hence *sar 'nan* for ‘easy day’, not **sar 'ar nan*. Likewise, this constraint requires the full form *arnan* in *tmr 'e ar.'nan 'leir* ‘a day of speaking’ rather than **tmr 'e nan 'leir*, since the penultimate syllable in the verse and the final odd syllable must be stressed.

For the purpose of this chapter, this is only a simplified explanation of the interaction between prosody, morphology, and semantics in the chants. The phenomena above, coupled with the unique lexical items, render the chants mutually unintelligible with contemporary, spoken Akha. This adds to the uniqueness of the grammar, since the thematic documentation approach allows for a grammar and dictionary based on a highly endangered genre of a language, one that most native speakers do not grasp. Another impact that this approach has had on the grammar-writing process is that all data revolve around religion—the chants and their translations in contemporary Akha. The distinctive patterns found in the chants are best described in comparison to their translations in contemporary Akha, and despite being focused on a single genre, the size of the corpus still allows for observing grammatical regularities in both registers. In conclusion, though the corpus focuses on an endangered genre, it still offers a wealth of information for a grammar.

6.2 COMMUNITY USES. As a fieldworker, I would like to think that this project will have a positive impact on the community. Unlike the descriptive products mentioned above, the corpus has proved to be of much more use and value. At the very least, the community now has access to audio recordings, transcriptions and translations of the verses in ‘Buffalo One’. It is now up to members in the community to choose how they wish to use these materials. To start, in December of 2010 MAPS organized a workshop in Shan State of Myanmar (Burma) to bring the shaman from Thailand to meet with their counterparts in Myanmar. Representatives from China came as well. This was a three-day training workshop on traditional religion and culture. All team members from the thematic documentation project were asked to share the project with the other participants. They discussed using the corpus

¹⁶ There are similar rules for compound nouns and other disyllabic words; however, this is beyond the scope of this chapter.

to promote the indigenous religion through learners' materials for a future generation of shaman apprentices, with the intention to produce textbooks with accompanying karaoke VCDs of the shaman song for distribution to villages where there are youth interested in learning the shaman's texts. This is also related to MAPS's first major project; the NGO recently received a grant from the US Embassy in Thailand in order to initiate a project on cultural literacy. This is a two-year long project aimed at producing instructional video and literacy materials explaining how to carry out cyclical festivals and rites. One team member, Lawq Gaw, has taken it upon himself to write the manual. The book is entirely in Akha, 165 pages long (Wang 2011). Another team member has finished editing the accompanying video guide on how to conduct a funeral ceremony in early 2012. Finally, drafts of the first three chants were produced in 2010. These books are meant to be used along with the audio recordings. Each contains a preface with instructions on how to use the book, and each verse in the audio recording is followed by a translation in contemporary Akha and explanatory footnotes where needed.

In summary, although the resulting grammar may not be of use to the community, the corpus that the grammar and dictionary are based on is proving to be useful due to its focus on an endangered genre. I also hope that the corpus described above will be useful to researchers in other fields. Common themes in the chants include creation, life, death, crops, animals, plants, migration routes and so on. For example, the dictionary contains over fifty names of bamboo species, classified according to folk taxonomy (which is reflected in the word-form). This information may be of use to botanists or ethnobotanists. Furthermore, the chants are all sung, which may be of interest to ethnomusicologists. Additionally, as with many religions, the numbers 3, 5, 7, 9, and 12 carry significant importance in the chants—an interesting concept that has received little attention in anthropology and is still not discussed in religious studies.

7. CONCLUSION. It is my hope that this chapter will open a dialogue on thematic documentation for language description with the intent of building theme-based corpora for thematic grammars focusing on endangered domains. As mentioned in section 1, a comprehensive documentation project includes not only corpus, but also a dictionary and grammar based on the corpus. Each field setting is different, therefore making it difficult to provide generalizations for carrying out fieldwork for language documentation and description projects. However, by controlling for topic when setting out for a language documentation project, fieldworkers can better produce corpora that are more useful to the community, and applicable to a wider audience.

REFERENCES

- Austin, Peter K. 2010. *Endangered Genres*. <http://www.paradisec.org.au/blog/2010/12/endangered-genres>. (05 May, 2011.)
- Bradley, David. 1979. *Proto-Loloish*. London: Curzon Press.
- Bradley, David. 1997. *What did they eat? Grain crops of the Burmic groups*. *Mon-Khmer Studies* 27.161-70.
- Brooks, Tim. 2011. *Endangered Alphabets*. (26 May, 2011.)

- Geusau, Leo A. von (ed.). 2002. *Oer Zar: life cycle of Zaqnyiq peoples, animals, and plants*. Chiang Mai, Thailand: unpublished ms.
- Haviland, John B. 2006. Documenting lexical knowledge. In Gippert, Jost, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*, 129-162. Berlin, New York: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel (eds), *Essentials of language documentation*, 1-30. Berlin & New York: Mouton de Gruyter.
- Kukeuwsakem, Yotesapong. Personal communications. (09 June, 2010.)
- Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com> (01 June, 2010.)
- Mosel, Ulrike. 2011a. Lexicography in endangered language communities. In Austin, Peter & Sallabank, Julia (eds.), *The handbook of endangered languages*, 337-353. Cambridge: Cambridge University Press.
- Mosel, Ulrike. 2011b. Morphosyntactic analysis in the field – a guide to the guides. In Nick Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*, 72-89. Oxford: Oxford University Press.
- SIL. 2006. *Dictionary Development Project*. <http://www.sil.org/computing/ddp/> (01 January, 2012.)
- Terrell, Jake, Argawv Byevtseirguq, Ardoq Ceiganr and Miqder Saeduquq. 2010. *Pirma dawq – Aqkaq dawq – Pa la dawq Dir xaq naq liq. Archaic Akha – Modern Akha – English dictionary*. Unpublished ms.
- Wang, Jianhua. personal communication. (28 May, 2011.)
- Wang, Jianhua (ed.). 2011. *Tsawrxir pavqkhawq gurtsov: A manual of Akha funeral ceremonies*. Chiang Rai, Thailand: Mekhong Akha Network for Peace and Sustainability (MAPS).
- World Oral Literature Project. 2009. *Definitions and understanding oral literature*. <http://www.oralliterature.org/about/oralliterature.html>. (26 May, 2011.)

jake.terrell@gmail.com

ABBREVIATIONS

AG	Akha gloss	GEN	genitive
CH	chant	NPST	nonpast
CLF	classifier	PTC	poetic syllable or non-lexical vocable
EG	English gloss	\$	full syllable, as in (C)V