

**Similar languages, different dictionaries:
A discussion of the Lamjung Yolmo and Kagate dictionary projects**

**Lauren Gawne
The University of Melbourne**

Abstract

This paper presents a discussion of dictionary projects for Lamjung Yolmo and Kagate, two Tibeto-Burman languages of Nepal. Although the languages could be considered dialects of the same language, and the tools used are the same, workflow choices take into account the context of each project. In this paper I give a background into the role of lexicography in language documentation and description, as well as a brief introduction to both languages and the people who speak them. I then detail the workflow for each project, and consider the different factors involved in each. For both of these projects well-designed work practices have been part of the production of lexicographic materials. I demonstrate that fieldwork-based dictionaries are not just useful outputs, but also important tools in a linguist's relationship with the communities they work with.

Keywords: Tibeto-Burman languages, Yolmo, Kagate, language documentation, workflow, lexicography

1. Introduction

Linguistic fieldwork has long been a fertile domain for the production of lexicographic materials, with the creation of a wordlist or dictionary often a key aim. Although this is the case, there is very little discussion of how the production of a dictionary fits into a larger documentation project. In this paper I will discuss the production of lexical outputs for projects on Lamjung Yolmo and Kagate, two closely related Tibeto-Burman languages of Nepal. This will provide an illustration of the fact that, just as no two fieldwork situations are the same, no two dictionary outputs will be the same, but having a considered and flexible workflow is of primary benefit. This paper will also demonstrate that dictionaries are not just end-product outputs, but the production of a dictionary is a task in itself, and one that can help enrich the linguist's relationship with the communities with whom they work.

I will begin by looking at the role of lexicography in fieldwork in section 2, and provide a brief introduction to Lamjung Yolmo and Kagate in section 3. I then discuss the tools used and articulate the work-flow for each project in section 4, before sharing general observations about how these work-flows related to the various unique factors present in both field experiences in section 6. Although these are the experiences faced by many fieldworkers, I hope that by articulating these experiences we can come to learn more successful ways of integrating dictionary outputs into larger documentation projects. This will be of benefit to communities, and our experiences working with them, as well as ensuring that our work reaches wider audiences.

It is worth briefly noting that the materials I am referring to as 'dictionaries' in this paper would more likely be referred to as wordlists within lexicography, as the entries provide translations into other languages, rather than detailed definitions in the languages themselves. As I will discuss below, I think it is respectful to these communities to refer to these materials as dictionaries, as this is the role they believe these materials provide, and the term 'dictionary' provides a status not conveyed by any other term.

2. Lexicography and fieldwork

The documentation and description of the world's linguistic diversity is a sub-discipline of

linguistics, and has its own practical and theoretical implications. Within this field the 'Boasian Trilogy' of descriptive grammar, interlinearised texts and a dictionary is often held to be the ideal aim for project outputs (Evans and Dench 2006: 10–16). Thus, basic lexicography is something that should be part of a descriptive linguist's skill set.

While this trio of outputs has been a long-standing aspiration, the way that we go about creating them has changed with the increasing use of computers and specialised software. The influence of theoretical perspectives relating to information management means that there is now a strong focus on persistent and accessible data (Thieberger and Berez 2012). Linguists are now able to make many more recordings thanks to increasingly affordable digital recorders with high volumes of storage, and are able to create rich corpora of texts that can be analysed with variety of tools. Some of these tools make the compilation of lexicographic information more flexible than it used to be when items were manually entered onto system cards.

While dictionary-making is one role of the linguist working on documentation and description, it should be noted that fieldwork lexicographic outputs are different to traditional lexicographic processes (Cheilliah and De Reuse 2011: 228). Some of these features of fieldwork lexicography will become apparent in the discussion of my experiences below, however it is worth drawing attention to them at this point. The first is that the types of lexicography undertaken as part of a wider project are often opportunistic. While a linguist may do some work specifically collecting lexical items, there are other types of data to collect as well, and so some lexical items may be added to the lexicon through use in narrative or exposition. This may mean that there is a leaning towards the kinds of items that appear in the genres being collected, or that there may be some items in the lexicon where only one sense is captured, and not another. Another point of difference is that those trained in language documentation and description may have more or less formal training in lexicography, and in all likelihood, the majority of their skills will be learnt on the job or draw from their knowledge of related sub-disciplines such as semantics and phonology. A final difference is that while the goal of lexicographers is to produce a dictionary, or equivalent output, the role of the linguistic in documentation projects is far broader in scope. Not only are there all three elements of the Boasian Trilogy, but there are the expectations of the community to be met, as well as researchers' funding bodies and institutions. While there have been many outstanding dictionaries created as part of documentation work, there are many other languages where very little lexical data has ever been published. In this paper I advocate for the focus to shift to making data more accessible, rather than those working on smaller language documentation projects feeling that their lexical data does not live up to the expectation of professional lexicography. I will demonstrate my preference for this model of working by discussing two different projects I have been involved with.

3. Lamjung Yolmo and Kagate: language and migration

3.1. Origins and migration

Lamjung Yolmo (ISO-639 SCP) and Kagate (ISO-639 SWY) are two Tibeto-Burman languages spoken in the lower hills of Nepal's Himalayan ranges. Both are of the Central Bodic family, and are related to Sherpa and Standard Tibetan. The main population of what is now known as 'Yolmo'¹ originally migrated from the Kyirong area in what is now Southwest Tibet two to three hundred years ago and settled in the Helambu and Melamchi valleys (Clarke 1980: 83, van Driem 2001: 864, Desjarlais 2003: 7). There are estimated to be between 10,000 and 50,000 speakers of Yolmo in this area

1 Yolmo is also referred to in the literature as Yohlmo, Hyolmo and Helambu Sherpa.

(Hari and Lama 2004: 702-703). Yolmo speakers are Tibetan Buddhist of the Nyingma school of Mahayana Buddhism, and have traditionally undertaken subsistence agriculture.

Both the Lamjung Yolmo and Kagate groups migrated away from this area around a century ago. As can be seen in the map in Figure 1, while the Kagate speakers moved east and settled in the Ramechhap district, the other population moved west and settled in the Lamjung district of Nepal. To date there is no evidence as to whether these groups left at the same time, or why they left. There is some possibility that they were not of high social standing. Both groups are occasionally named as Tamang in earlier references (Höhlig and Hari 1976: 1, von Fürer-Haimendorf, unpublished fieldnotes Nepal 1957, no. 12, p. 278), which is a group of lower social standing in the original area (Clarke 1980). Neither group speak Tamang language, however they appear to be linked to a lower social status than other Yolmo speakers. There are also a group of Yolmo speakers in the Ilam district in far-east Nepal, as can be seen in Figure 1. This as yet undocumented dialect has been discussed in Thokar (2009), and further demonstrates that migration away from the traditional Yolmo villages has occurred for several groups.

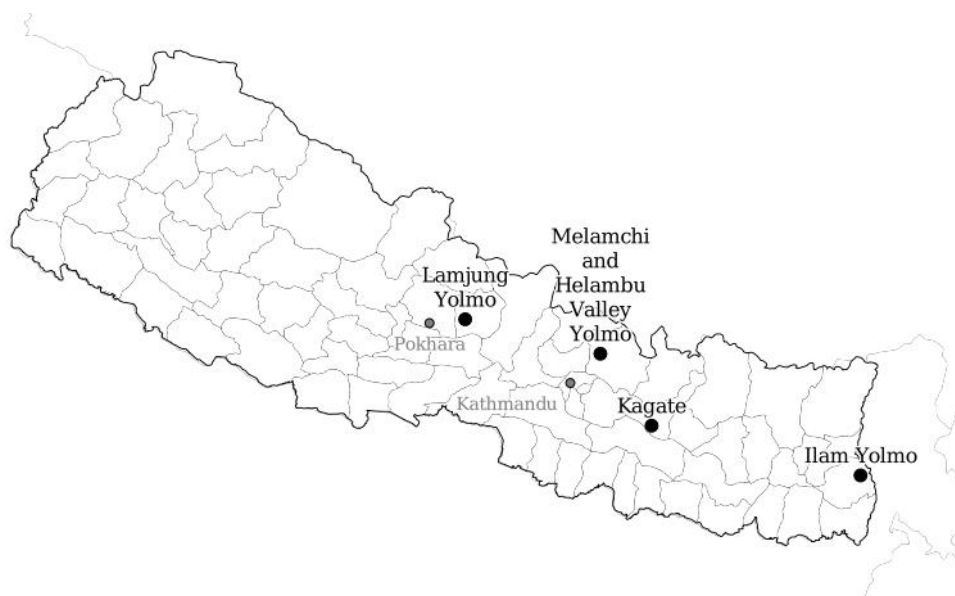


Figure 1: Map of Nepal showing the location of Yolmo spoken in the Melamchi and Helambu Valleys and the groups of Lamjung Yolmo and Kagate speakers who migrated away.

Neither the Lamjung Yolmo or Kagate groups of speakers now have any connection to their ancestral homes in the Melamchi and Helambu Valleys. The Kagate are no longer known as 'Yolmo' but have taken on the name 'Kagate'. The Nepali term for paper is *kagate*, and both the Kagate of Ramechhap and the Yolmo of Lamjung have been referred to exonymously as Kagate. The profession of paper maker is considered to be a low caste occupation in the strict Hindu caste system that existed in Nepal. Whether this was an occupation they held as lower Yolmo speakers before leaving the Melamchi and Helambu Valleys, or an occupation both groups adopted independently is something to still be investigated. The Yolmo of Lamjung find the name Kagate to be highly pejorative, based on its lower-caste history, and therefore have identified their language with the main group of Yolmo. On the other hand, the Kagate speakers of Ramechhap that I have spoken to are all proud to identify as Kagate. This may be an unexpected flow-on effect of language documentation work in the 1970s.

3.2 Languages or dialects

Given their common history, we would expect a strong level of similarity between these groups. Ethnologue (Lewis 2009) states on the Helambu Sherpa information page that the language is not mutually intelligible with Kagate. However, Hari (2010: 1), who has worked extensively with both languages, argues that ‘to quite a large extent they are mutually intelligible dialects.’ I have also introduced speakers of Kagate and Lamjung Yolmo to each other, and they have been able to converse easily, although each noticed that the other's speech was observably different to their own.

Gawne (2010) presented a small-scale survey of the lexical similarity of these dialect groups following the method outlined in Blair (1990). The lexical similarity of the three languages was high, with Melamchi/Helambu Yolmo and Kagate having at least a 79% lexical similarity, Melamchi/Helambu Yolmo and Lamjung Yolmo having an 85% similarity and Lamjung Yolmo and Kagate sharing an 88% lexical similarity. Further work in this area, plus the experience of observing speakers interact, indicates that the lexical similarity is likely to be even higher. This is very high when we compare the lexical relationship to other Tibeto-Burman languages identified as sharing similarities with Yolmo; according to Ethnologue (Lewis 2009) Yolmo has a lexical similarity of 65% with Lhasa/Standard Tibetan and 61% with Sherpa.

Therefore, there is no lexically motivated reason why speakers might want to have separate dictionaries for their own dialects. There are some variations in the orthographic standards used by speakers of the different varieties – with Kagate speakers preferring not to differentiate long and short vowels, and Lamjung Yolmo speakers preferring to – however there is no interest in moving towards a standardised orthography, and no consensus about what form it would take. It would be perfectly feasible to create a dictionary that captured key lexical similarities, as well as dialectal variation, however as discussed below, speakers have their own reasons for not wanting this that go beyond linguistic factors.

3.3 Previous work

Extensive documentation work has been carried out on the main variety of Yolmo over the last twenty years by Annemarie Hari. This has resulted in the publication of a large and comprehensive Yolmo-English-Nepali dictionary (Hari and Lama 2004) as well as a grammar (Hari 2010). Even though this dictionary is impressive, speakers of Lamjung Yolmo and Kagate have not really embraced it as representing their language. For Kagate, the name alone appears to be enough to have motivated this, but for both groups the lack of connection with this main group of speakers means that they do not identify a dictionary produced by the main Yolmo group as representative of the language that they speak, even though many of the words are recognisable.

Although Kagate (spoken in Ramechhap, as discussed above) has a much smaller speaker population than Yolmo (1,500 compared to at least 10,000) it has, until recently, received much more attention from linguistic researchers. Höhlig and Hari (1976) produced a comprehensive phonemic summary of the language, and Höhlig (1978) wrote a paper on speaker orientation. Höhlig and Hari also created a typewritten wordlist, with Kagate and English, as well some entries with Nepali and German lexical information. A copy of this wordlist was given to their main language speaker.

My work on Lamjung Yolmo constituted the first known documentation of this variety. Therefore, even though Lamjung Yolmo and Kagate are linguistically similar, these projects are

different in regards to their materials available throughout the process.

The Lamjung Yolmo dictionary was created at the request of several of my main consultants, as part of my obligation in exchange for their assistance in the wider documentation project. During my time in Nepal working Lamjung Yolmo I made contact with Kagate speakers. I offered to help them update the existing materials that they had from the earlier documentation work in the 1970s.

4. Dictionary tools and workflows

In this section I will list the tools and work processes used for each of these projects. Although both projects used the same set of software, the way this software was used differed for both projects.

4.1 Tools used

Both dictionaries were made using the same basic tools. Lexical databases were created in Toolbox (Buseman and Buseman 2009) before being exported using the Multi-Dictionary Formatter (MDF) function. From there a small amount of editing was done in a word processor, before print and digital copies were made.

These tools are useful in that they are open-source and there is still community support and documentation available. There is support for non-Roman alphabets, which is useful as the preferred orthography is Devanagari. Toolbox also has a degree of interoperability with ELAN (Hellwig, Van Uytvanck and Hulsbosch 2009), which means that it is useful as part of a larger project that also involves transcribing and interlinearising audio recordings. This means that building up the lexicon can also be part of the process of interlinearising texts, thus showing the usefulness of software in helping researchers attempt the Boasian ideal of multiple outputs from the same project, as part of a single workflow.

Although Toolbox was central to the storage and formatting of the dictionary for each of these projects, the point at which it was used has been different. I will now go through the Lamjung Yolmo and Kagate projects in turn to describe the dictionary workflow and its role in the larger documentation project.

4.2 Producing the Yolmo dictionary

Work on Lamjung Yolmo was undertaken for my Doctoral dissertation, which was the main output for the project. I agreed that in exchange for the assistance of Lamjung Yolmo speakers I would produce a dictionary of their language. Therefore, even though it was not the main focus of the project, I knew from early in the process that I would have to create a publishable version of the lexical information I had been collecting.

The lexicon was built up over two fieldtrips of 8 months in total. Many basic lexical items were collected as part of word-list elicitation, especially early in the project. Other lexical items were added to the database when texts were being transcribed, which often proved to be a rich source of verbal and adjectival data. I also had some success with providing one of my regular teachers and her literate daughter with an opportunity to gather their own data. I gave them a

blank exercise book with headings at the top of pages for basic semantic categories. In this way some 50 additional lexical items were collected. While not a high number, it did allow the participants to feel ownership of the project. It also helped them understand the type of work I was trying to do, as well as understanding some of the challenges of lexicography.

The dictionary needed to be finalised so that I could return with it on the final fieldtrip for the project. By that point there were around 1200 entries in the lexicon. While this is not a particularly large number for a lexicographic project, it represented a relatively concerted effort for an output that was secondary to the main aim of the project. Exporting and formatting the dictionary was relatively expedient, in comparison to the data collection phase. I had included both Roman and Devanagari forms of the words at the input stage, so was able to tailor the output to the Devanagari-reading Yolmo audience and sections with Nepali and English headwords translating to Lamjung Yolmo were made possible by the 'reverse' field in Toolbox, which I had been populating as part of adding new entries since early in the documentation process.

So as to make the dictionary as useful as possible, I decided to choose a print-on-demand model, so that more copies could be made in the future. Also, although the primary audience for the dictionary is the community of Lamjung Yolmo speakers, this does not mean that it would not be relevant to others, and so a print-on-demand model makes the data available to others as well. I also made a PDF version available through the World Oral Literature Project (www.oralliterature.org). I was able to take 40 copies with me thanks to the funding from Rachel Nordlinger and Nicholas Thieberger's ARC Discovery grant "Doing great things with small languages" (DP0984419).

While the reaction to the dictionary was mostly positive (and occasionally ambivalent) several reactions are worth noting. The first is the problems encountered with the initial blurb on the outside back cover. This first version of the blurb stated that Lamjung Yolmo is "related to Helambu Sherpa and Kagate". As I demonstrated in §3.2 above, this is not at all a contested linguistic fact. Lamjung Yolmo speakers, however, were offended that their language was being equated with something that, for them, represented strongly pejorative past. By having a dictionary for Lamjung Yolmo they were asserting their Yolmo identity, and even passing mention of Kagate was met with strong disapproval. All subsequent printings have replaced Kagate and Kyirong, and some very low-tech paper-and-glue editing was employed to make the copies I took over with me acceptable for distribution. While stressful at the time, it did allow for a much deeper discussion about the relationship between social status, language and name than I'd had with people previously, and gave me new insight to the attitudes of speakers. One person who had not engaged with my work before went one step further, and went through the whole book and returned it to me with his editing notes. While many of his notes related to a difference of opinion regarding orthography, there were interesting notes regarding many items. By producing the dictionary I was able to draw the interest of a much wider group of speakers, as well as demonstrating to them my commitment to their language and reliability in fulfilling my promises.

Although the dictionary is completed, there is still much work to be done in regards to the Lamjung Yolmo lexicon. The dictionary produced was limited by the fact that it developed as my skills at Toolbox were also developing. With the knowledge I have now, I could have designed a better lexical database from the outset, however doing things like adding example sentences and additional notes post hoc is time-consuming. Also the print-on-demand service used for the initial run has now closed. As this is a relatively new printing model it's not surprising that there is some instability in the services provided, however given that one of the primary aims for this project is

to make the data accessible to all it would be good to ensure people can buy print copies in the future.² There are also plans to make an online version of the dictionary. Thanks to the fact that it exists as a lexical database, and not just a printed book, this should be a relatively easy undertaking.

4.3 Producing the Kagate dictionary

The Kagate dictionary, in its current format, has around 1700 entries, and was also formatted using Toolbox and MDF. Beyond the similarities in tools used, there are key differences between the two projects. I first met speakers of Kagate when I was given their contact details by Annemarie Hari in 2009. This included the speaker with whom Hari and Höhlig had worked in the 1970s. We stayed in touch, although my focus on my PhD topic prevented me from working with them at the time. In 2010 we agreed that there was a pressing need to ensure that a type-written single-copy wordlist left with these Kagate speakers by Hari and Höhlig in the 1970s was preserved in some way.

We made high quality scans, which at least ensured a back-up version of the dictionary now existed. These digital scans were then passed through Adobe Acrobat 9's optical character recognition (OCR) filter by my colleague Amos Teo. From there, some characters needed regularising, and some alignment work had to be done. We now had a searchable text version of the data. Utilising some of the formatting elements of the original text, we were able to create all of the backslash codes needed to allow the file to be read by Toolbox. For example, any Nepali definitions in the wordlist were fronted with *Np:*, so I was able to search for this string and replace it with *\gn*, which meant Toolbox recognised it as a 'national language gloss' (i.e. the Nepali translation). This process is much more expedient than manually re-entering the data. It also means that now we can create a range of outputs, the first of which was a clean, reformatted version of the original wordlist to be returned to speakers. Now multiple copies can be made and distributed. This is the current status of the Kagate project, and while it is certainly in a better shape than it was before, there is still much to be done.

Toolbox featured very late in the process for the creation of this dictionary – and at this point has largely been used for its convenient formatting properties. There are many future plans for this data. Now that it is in an accessible database, it can form the basis of a new documentation process. As it was typewritten, there was no Devanagari in the original version, which is the main script of literacy for these speakers. We hope to enrich the current database with Devanagari, as well as example sentences. There is also no lexical tone value, which needs to be captured. This lexicon will be enriched as it is used for interlinearising texts. Several of the younger urban speakers with computer literacy also now have access to a laptop (and have had basic training in Toolbox and lexical input). It is hoped that much like the blank exercise book given to Yolmo speakers, this will be a tool that will be used to engage with the project and give speakers a sense of ownership of the task ahead.

5. Discussion: Project observations

Above I explain the process used for both of the projects. In this section I will discuss some of the observations I have made during each of these projects. These observations have been grouped into five main themes. The first is the expectations of the two groups of people I have

² The Lamjung Yolmo-Nepali-English dictionary is once again available as a print-on-demand book through Create Space. This Amazon subsidiary is hopefully large and stable enough that this will be a more sustainable option.

worked with. Second, I wish to reflect on my own expectations regarding each project. I then turn to a number of constraints faced during these projects, namely those pertaining to time, budget and data collection opportunity.

With both of these projects I have tried to work with community expectations as close as possible. This has largely been achievable because the outputs expected by community and the outputs required of linguists working within academia can be developed simultaneously. The texts that formed the core of the dissertation analysis in Gawne (2013) were the same texts that produced many of the lexical items in Gawne (2011). The dictionary itself was the work of a week, while the building of the lexical database occurred over 8 months of intensive fieldwork.

There is some discussion about whether we should be trying to design dictionary outputs from fieldwork for academics or communities (Hinton and Weigel 2002). This discussion is somewhat less relevant if we concern ourselves with building quality lexical databases, which can then be used to build outputs for either linguists or communities with (relatively) little work. Although there is now a Lamjung Yolmo dictionary that is currently created with the needs of the community in mind, there is no reason why I could not also produce something that would be much more suited to the needs of academics, with more grammatical information presented, and the linguistic notes I have added over time.

That each group wanted their own record gives an interesting insight into the social standing between what are, in reality, closely related languages. Not only did I find that planning both dictionary projects (and delivering them) allowed us to discuss specific lexical items, the process allowed for much interesting discussion about language names, language status and attitude to language use. It is these discussions that have led me to continue to refer to these outputs as 'dictionaries' even though they more closely resemble wordlists. They are treated as dictionaries by the communities that requested them, and to refer to them as such is not important for the sake of the documents themselves, but what they represent to the communities who were interested in their development.

It also means that speakers are more interested in elements of the project that they may otherwise find esoteric. For Lamjung Yolmo it meant that I was able to collect a greater diversity of narrative, because speakers were interested in the possibility that there may be novel lexical items. For both languages it has also made my role as linguist clearer, allowing me to ask questions about the nature of household items in ways that may have otherwise been interactionally opaque to participants. This has led to another useful benefit of the production of these resources, especially the Lamjung Yolmo dictionary, in that it has shown people that I am a dedicated student of the language, and a willing participant in their community. Being able to return with some of the fruits of our labours has led to some interactions that were as personally heartening as they were professionally rewarding.

In regards to my own expectations, these projects have developed as my technical competence and knowledge of the languages developed. Indeed, having further developed my confidence at using Toolbox, and familiarity with the language, I feel that for any future project I will include more example sentences and cross-referencing from the earlier stages of development. One other thing that shifted during the course of these two projects has been my attitudes towards data access. Although neither project is perfect, or anywhere near detailed enough to be called a fully representative dictionary of either language, I would prefer that these less-than-perfect records exist in ways that are accessible to others. This 'letting go' has been difficult, but

the reception has been largely positive, and has made me feel more confidence that unlike the first iteration of the Kagate dictionary, when photocopying was much less accessible, this version will not end up as a single copy on someone's shelf.

There were a number of constraints that shaped both of these projects that I think are worth discussing. The first is the nature of time constraints on both projects. In regards to the Kagate dictionary, this has been, until now, a side project. This is not just the case for me, but for the speakers of the language who I work with, who all have ongoing employment. Therefore, harnessing existing materials has allowed us to start a new period of documentation with a good foundation. In regards to the Lamjung Yolmo project, the need to focus on the documentation of grammar and collection of texts meant that there was less time to undertake concerted lexical research. There are still under-defined items in the lexicon, for example *yálo tséema* and *kòlo tséema* are both glossed as 'leafy greens', but are different varieties. There was also a small window of time available to take the lexicon database and produce an output that resembled a dictionary. This was one time where the time constraint was not so problematic, as the data had been kept in a format that allowed for relatively easy export.

Another issue has been the financial constraints involved in undertaking a project of this nature. Even though communities are very eager to have dictionaries and other printed materials, and the production of these materials improves their participation and positive attitude towards language documentation projects, funding bodies are often reluctant to provide funding for the printing of dictionaries. Even though printing costs for a print-on-demand model can be quite low, I was very fortunate with the Lamjung Yolmo dictionary to have the support of colleagues who assisted me with these costs.

One obvious difference between these two projects has been the way that the data has been entered in Toolbox. While the Lamjung Yolmo dictionary was an aggregate of all of the lexical items that had been observed until that time in the documentation project, the Kagate dictionary involved the reformatting of existing data. Both dictionaries can be considered as opportunistic, in that they have been largely been created from data that had other functions as well. With Kagate speakers, the opportunities for data collection have so far been few. This means that utilising available resources has been an important part of beginning this project. For Yolmo, while there were many data-collection opportunities, there were other tasks that needed to be prioritised. Collecting lexical data was one element of these tasks, and allowed me to continue to slowly develop a lexicon of the language, and that lexicon will continue to be enriched during any future work.

6. Conclusion

Although Kagate and Lamjung Yolmo are closely related enough as to be mutually intelligible, speakers of each variety have been very keen to develop their own lexical resources, which have lead to two very different project experiences. Even though both projects have been developed using Toolbox and MDF, they have taken shape differently. There are similar constraints regarding budget, time and elicitation opportunities, which are shared with many other fieldwork lexicography projects. One common element is that the dictionary, even though it is only one part of the language documentation and description process, has helped positively shape my experience in both communities.

Although there is often very little professional recognition for fieldwork outputs of this type, I still believe that they are important. This is not only because of the ethical obligation towards the communities that we work with, but also the professional obligation to make data readily available to the wider academic community. The Boasian Trilogy with dictionaries as a key element is still relevant today. With resources that make it possible to work on developing lexicographic information at the same time as other outputs, this task is a viable part of fieldwork, although how it fits into the process may differ from project to project. Having a considered workflow that fits these expectations, and the expectations of speakers is one way to ensure that lexicon collection is maximally useful for all concerned.

Acknowledgements

Thankyou to participants at the Australex annual conference 2013 for the feedback and thoughtful discussion. Thankyou to the Yolmo and Kagate speakers who made these projects possible. The Lamjung Yolmo dictionary was funded in part by Rachel Nordlinger and Nick Thieberger's ARC Discovery grant "Doing great things with small languages" (DP0984419), and supported by the World Oral Literacy Project. The Kagate dictionary has been sponsored by the Awesome Foundation (Ottawa chapter) and a Stack Exchange fellowship. Thanks to Amos Teo for his OCR skills.

References

- Blair, Frank. (1990). *Survey on a shoestring*. Arlington: SIL International and The University of Texas.
- Buseman, Karen, & Buseman, Alan. (2009). Toolbox (version 1.5.5) [Computer Program]. Retrieved 30st April, 2009, from <http://www.sil.org/computing/toolbox/downloads.htm>
- Chelliah, Shobhana L., & De Reuse, Willem J. (2011). *Handbook of descriptive linguistic fieldwork*. London: Springer.
- Clarke, Graham E. (1980). Lama and Tamang in Yolmo. In M. Aris & A. S. S. Kyi (Eds.), *Tibetan Studies in honor of Hugh Richardson* (pp. 79-86). Warminster: Aris and Phillips.
- Desjarlais, Robert R. (2003). *Sensory biographies : lives and deaths among Nepal's Yolmo Buddhists*. Berkeley, Calif.: University of California Press.
- Driem, George van. (2001). *Languages of the Himalayas : an ethnolinguistic handbook of the greater Himalayan Region : containing an introduction to the symbiotic theory of language*. Leiden ; New York: Brill.
- Evans, Nicholas, & Dench, Alan Charles. (2006). Introduction: Catching language. In F. K. Ameka, A. C. Dench & N. Evans (Eds.), *Catching language : the standing challenge of grammar writing* (167 ed., pp. 1-40). Berlin: Mouton de Gruyter.
- Gawne, Lauren. (2010). Lamjung Yolmo: a dialect of Yolmo, also known as Helambu Sherpa. *Nepalese Linguistics*, 25, 34-41.
- Gawne, Lauren. (2011a). *Lamjung Yolmo - Nepali - English dictionary*. Melbourne: Custom Book Centre, The University of Melbourne.
- Gawne, Lauren. (2013). *Lamjung Yolmo copulas in use: Evidentiality, reported speech and questions*. (PhD), The University of Melbourne, Melbourne.
- Hari, Anna Maria. (2010). *Yohlmo Sketch Grammar*. Kathmandu: Ekta books.
- Hari, Anna Maria, & Lama, Chhegu. (2004). *Hyolmo-Nepālī-Aṅgrejī śabdakośa = Yohlmo - Nepali - English dictionary*. Kathmandu: Central Dept. of Linguistics, Tribhuvan University.
- Hinton, Leanne, & Weigel, William F. (2002). A dictionary for whom? Tensions between academic and nonacademic functions of bilingual dictionaries. In W. Frawley, K. C. Hill & P. Munro (Eds.), *Making dictionaries: preserving indigenous languages of the Americas* (pp. 155-170). Berkeley; London: University of California Press.

- Höhlig, Monika. (1978). Speaker orientation in Syuwa (Kagate). In J. E. Grimes (Ed.), *Papers on discourse* (Vol. 50, pp. 19-24). Kathmandu: Summer Institute of Linguistics.
- Höhlig, Monika, & Hari, Anna Maria. (1976). *Kagate phonemic summary*. Kathmandu: Summer Institute of Linguistics Institute of Nepal and Asian Studies.
- Lewis, M. Paul. (2009). Ethnologue: Languages of the World, Sixteenth edition. 16. Retrieved 11 February 2009, 2009, from <http://www.ethnologue.com>
- Thieberger, Nicholas, & Berez, Andrea L. . (2012). Linguistic Data Management. In N. Thieberger (Ed.), *The Oxford handbook of linguistic fieldwork* (pp. 90-118). Oxford: Oxford University Press.
- Thokar, Rajendra. (2009). *Linguistic fieldwork in Jhapa and Ilam districts*. Paper presented at the Linguistics Society of Nepal Annual Conference, Kathmandu, Nepal.