

# 3

## Training for Language Documentation: Experiences at the School of Oriental and African Studies

Peter K. Austin

*School of Oriental and African Studies, University of London*

Since 2003 the Endangered Languages Project at SOAS has been involved in various types of training for documentation of endangered languages, ranging from one-day workshops through to MA and PhD post-graduate degree programs. The training events have been attended by specialists, research grantees, students, and members of the general public, and have covered a wide range of topics and involved delivery in a range of contexts and delivery modes, including hands-on practical sessions and e-learning in the Blackboard framework. We have covered both theory and practice of language documentation and endangered language support, including the development of multimedia and curriculum materials for language teaching, some of it experimental and, we think, quite innovative. In this paper I discuss some of our experiences in developing and running these training workshops and courses, reporting on the models, and successes (and failures) over the past three and a half years. My goal is to share our accumulated knowledge and experience with others with similar interests, and in doing so to advance our understanding of the possibilities for language documentation training.

**1. INTRODUCTION.**<sup>1</sup> The last ten years has seen the emergence of the new field of Documentary Linguistics as an area of research that deals with the principles and practices of documenting languages, particularly endangered languages. Correspondingly, a need has developed for training both practicing linguists and a new generation of students in the theory and methods of language documentation. This paper is an overview of the training programs we have set up at the School of Oriental and African Studies (SOAS), University of London, to address this need. Section 2 outlines documentary linguistics and language documentation as generally conceived, and Section 3 identifies the range of skills required to undertake language documentation. Section 4 introduces the Hans Rausing Endangered Languages Project (HRELP), and in Section 5 I discuss the training models we have developed at SOAS to introduce and extend language documentation knowledge and skills for post-graduate students and established linguists who have received a grant from HRELP. In Section 6 I give an assessment of some of the successes and challenges we have seen over

---

<sup>1</sup> This paper was written for the International Conference on Austronesian Endangered Language Documentation, Providence University, Taiwan 5-7 June 2007. The ideas and materials presented here have been discussed in some detail with Oliver Bond, Lenore Grenoble, Colette Grinevald, Anthony Jukes, Friederike Luepke, Sophie Manus, Robert Munro, and David Nathan. I am grateful for feedback on an earlier version from two anonymous referees. None of these people can be held responsible for any errors or shortcomings remaining.

the five years HRELP has been operating. Section 7 presents some conclusions from our experience that may be of interest and value to others wishing to establish training courses for language documentation.

**2. DOCUMENTARY LINGUISTICS AND LANGUAGE DOCUMENTATION.** Documentary linguistics is a newly emerging field of linguistics that is “concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties” (Gippert, Himmelmann and Mosel 2006:v). Documentary linguistics has developed over the last decade in large part in response to the urgent need to make an enduring record of the world’s many endangered languages and to support speakers of these languages in their desire to maintain them (Whalen 2003, Austin 2007). It is also fueled by developments in information, communication and media technologies which make documentation and the preservation and dissemination of language data possible in ways which could not previously be envisioned. In addition it essentially also concerns itself with the roles of language speakers in documentary projects and their rights and needs in ways not previously considered within linguistics (see Thieberger and Musgrave 2007).

Himmelmann (2006:15) identifies several important new features of documentary linguistics:

- Focus on primary data – language documentation concerns the collection and analysis of an array of primary language data to be made available for a wide range of users;
- Explicit concern for accountability – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected;
- Concern for long-term storage and preservation of primary data – language documentation includes a focus on archiving in order to ensure that documentary materials are made available to potential users into the distant future;
- Work in interdisciplinary teams – documentation requires input and expertise from a range of disciplines and is not restricted to linguistics alone;
- Close cooperation with and direct involvement of the speech community – language documentation requires active and collaborative work with community members both as producers of language materials and as co-researchers.

I use the term language documentation to refer to the activities carried out by researchers and communities engaged in work that adopts a documentary linguistic approach. The historical genesis of the field of documentary linguistics has meant that the term ‘language documentation’ is sometimes used loosely, to refer to any kind of language record, but documentary linguistics uses it in a more specific way, to refer to an activity with much larger and more specific goals. In particular, language documentation strives “to provide a comprehensive record of the linguistic practices characteristic of a given speech community” (Himmelmann 1998:166). Language documentation differs fundamentally and critically

from language description. Language documentation seeks to record the linguistic practices and traditions of a speech community, along with speakers' metalinguistic knowledge of those practices and traditions. This includes systematic recording, transcription, translation and analysis of the broadest possible variety of spoken (and written) language samples collected within their appropriate social and cultural context (Austin 2006, HRELP 2006). Analysis within language documentation is aimed at making the records, or rather the language data recorded, accessible to a broad range of potential users. This group includes not only linguists but also community members, who may not have first-hand knowledge of the documented language. The record is thus intended for posterity, and so some level of analysis is required, in particular glossing and translation into one or more languages of wider communication (see Evans and Sasse 2007 for some of the challenges that entails), and systematic recording of metadata to make the archived document(s) findable and usable (Nathan and Austin 2004).

I take the core of a language documentation project to be the creation of a corpus of audio and/or video materials with time-aligned transcription, multi-tier annotation, translation into a language of wider communication, and relevant metadata on context and use of the materials. Woodbury (2003) argues that the corpus will ideally be large, cover a diverse range of genres and contexts, be expandable, opportunistic, portable, transparent, ethical and preservable. As a result, documentation is increasingly done by teams, including community members, rather than 'lone wolf linguists;' both the technical skills and the amount of time required to create this corpus make it difficult for a single linguist, working alone in the field, to achieve.

Language documentation typically begins with the development of a project to work with a speech community on a language and can be seen as progressing through a series of stages, some of which are carried out in parallel:

- Project conceptualization and design
- Establishment of field site, including negotiation of permissions
- Funding application
- Data collection and processing
- Creation of outputs
- Evaluation and reporting

The following stages in the data collection and processing phase can be recognized (Austin 2006):

1. *recording* – of media (audio, video, image) and text
2. *capture* – moving analogue materials to the digital domain
3. *analysis* – transcription, translation, annotation, and notation of metadata

4. *archiving* – creating archival objects, and assigning access and usage rights

5. *mobilization* – publication, and distribution of the materials in various forms

Language description typically involves the production of grammars, dictionaries, and collections of texts<sup>2</sup>. In contrast, the primary goal of language documentation is the development of a corpus which is representative of a wide range of discourse types (Austin 2006, Woodbury 2003, Himmelmann 1998). Although description relies on documentation (and documentation essentially includes descriptive aspects such as annotation), it involves analysis of a different order: description provides an understanding of language at a more abstract level, as a system of elements, rules, constructions and so on (see again Himmelmann 1998, 2002:48). Description and analysis are contingent by-products of documentation and will change and develop over time as research progresses (Woodbury 2003, Austin and Grenoble 2007). Such works can be valued by speech communities and provide important input into processes of language maintenance and revitalization, however, the primary audience for these products is typically linguists, and sometimes they are written in frameworks accessible only to trained linguists. Such products can also become unusable as linguistic theoretical models come into and go out of fashion. Thus many of the grammars written in the 1970's and 1980's in Tagmemic or Transformational Grammar frameworks are now extremely difficult to use. Language documentation focuses specifically on providing a preservable and transparent corpus of analysed materials on a language that is well structured and designed for access by non-linguists as well.

**3. SKILLS FOR LANGUAGE DOCUMENTATION.** Language documentation requires knowledge and application of a range of skills, including those traditionally associated with fieldwork and language description, as well as skills in the application of information, communications and media technologies (Munro 2005) and applied ethics. Increasingly also, documenters are expected to have knowledge and skills typically associated with areas of applied linguistics, such as orthography development, lexicography, translation, pedagogy and curriculum design, multimedia, language policy and needs assessment, and advocacy. The need for these skills arises from the desires and expectations of the language communities and the multidisciplinary orientations of the work.

I propose the following is an (incomplete) list of documenter skills that researchers should have some exposure to and competence in:

- *Project conception, design and management* – familiarity with documentation theory, applied ethics, intellectual property rights and socio-cultural issues, stakeholder communication
- *Grant application writing*

---

<sup>2</sup> Historically, and in some cases currently, some linguists use the term 'language documentation' to refer to what we are calling 'language description'. I attempt to be consistent in my usage of the two terms.

- *Media management* – recording techniques, field methods, data transfer, backup
- *Data and metadata management* – data and metadata representation (XML, relational database models), transcription, linguistic analysis (phonetics, phonology, morphology, syntax, semantics) and annotation, use of linguistic software tools (Transcriber, Shoebox/Toolbox, ELAN, IMDI), data integrity and sustainability (Bird and Simons 2003), workflow design and management
- *Mobilization* – familiarity with applied linguistics concepts (orthography design, lexicography, curriculum development, policy formation, revitalization), publication skills, multimedia design and implementation (Csato and Nathan 2003, Nathan 2006, Nathan and Csato 2006)
- *Team-based research* – skills sharing and transfer, capacity development
- *Reporting* – presentation, writing and communication skills

It is highly unlikely that any single researcher will have all this knowledge and skills at a sufficient level to be able to apply them all in carrying out a project. Indeed, as Walcott (1999:73) says of ethnography:

“instead of envisioning the ethnographer as Superman or Wonder Woman, one must recognize that it is the scope of the ethnographic question that must be pared to what one individual, or a researcher working with a colleague or small research team, can accomplish in a limited amount of time. The range of fieldwork techniques that can be employed must be pared as well. Of course, in an ideal world, every researcher would be sufficiently talented to be able to summon from a vast personal repertoire whatever combination of techniques seems appropriate to addressing the issue at hand.”

In other words, projects need to be realistic in terms of what can be achieved in the time available, and researchers must ensure that they acquire the skills they need for that particular project. Indeed, Walcott (ibid) argues that:

“if you, as sole or principal investigator, really did have all the skills of social research at your command – computer skills, language skills, statistical skills, survey techniques, ability to work with experimental and quasi-experimental design – let alone all the observer and interviewer and interpersonal skills an ethnographer is likely to need, why would you invest your time plodding along with ethnography?”

We just need to substitute ‘language documentation’ for ‘ethnography’ here to make his advice applicable to our current concerns.

**4. THE HANS RAUSING ENDANGERED LANGUAGES PROJECT.** The Hans Raising Endangered Languages Project (HRELP) was established with a commitment of £20 million from Arcadia (formerly the Lisbet Raising Charitable Fund) to document as many endangered languages as possible and to encourage the development of relevant skills across the world. It has the following three components:

1. *Endangered Languages Documentation Program (ELDP)* – will provide approximately £15 million over an 8-10 year period in competitive research grants to encourage the development of linguistic fieldwork in endangered languages (especially by younger scholars) and to support documentation of as many threatened languages as possible. ELDP offers five types of grants, and is governed by an international selection panel chaired by Prof. Graham Furniss of the School of Oriental and African Studies (SOAS); the ELDP grants administration is managed by SOAS.

2. *Endangered Languages Academic Program (ELAP)* – aims at training the next generation of language documenters, it offers students an array of opportunities: an MA in Language Documentation and Description, a PhD in Field Linguistics, and post-doctoral fellowships at SOAS. We also offer a comprehensive program of public lectures, seminars, workshops and training courses. Prof. Peter K. Austin, Märit Raising Chair in Field Linguistics, is Director of ELAP.

3. *Endangered Languages Archive (ELAR)* – is building up a large collection of endangered languages documentation resources, and also supports training, development of methodologies, and other technical aspects of language documentation and archiving. David Nathan is Director of ELAR.

HRELP has involvement in 18 research projects concerned with endangered Austronesian languages, as set out in the following Table<sup>3</sup>. All projects except one are funded by ELDP and some involve post-graduate students or post-doctoral researchers who are being trained by ELAP; all other grantees have attended training courses at SOAS (see 5.2 below). All the data collected by these projects is being archived at ELAR.

TABLE 1. HRELP: Endangered Austronesian Languages Research Projects

Location	Language	Researcher & Institution <sup>1</sup>	Notes
Taiwan	Yami	Der-Hwa Victoria Rau, Providence University, Taiwan	grant 2005-07
Philippines	Palawan	Charles Macdonald, CNRS, France	grant 2006-07
Micronesia	Pingilapese	Ryoko Hattori, University of Hawai'i at Mānoa, USA	grant 2006-07

<sup>3</sup> For further details see <http://www.hrelp.org/grants/projects/>

Indonesia	Toratán	Anthony Jukes, SOAS	grant 2005-07, ELAP post-doc
	Helong	John Bowden, ANU, Australia	grant 2007-10
	Rongga	Wayan Arka, ANU, Australia	grant 2004-06
	Allang Amahai Tulehu	Margaret Florey, Monash University, Australia	grant 2003-05
East Timor	Maku'a	Aone van Engelenhoven, Leiden University, Netherlands	grant 2003
Malaysia	Eastern Penan	Peter Sercombe, Northumbria University, UK	grant 2003-04
New Caledonia	Numèè	Sophie Rendina, SOAS	grant 2007-09, ELAP PhD student
Vanuatu	Mafea	Valérie Guérin, University of Hawai'i at Mānoa, USA	grant 2005-07
	Vurès	Catriona Malau (nee Hyslop), La Trobe University, Australia	grant 2004-06
	Biriebo	Peter Budd, SOAS	grant 2005-07, ELAP PhD student
	Seke	Kay Johnson, SOAS	grant 2007-09, ELAP PhD student
	Neverver	Julie Barbour, Waikato University, New Zealand	grant 2004-07
	Qatareu	Hans Schmidt, Hamburg University, Germany	grant 2004-06
Solomon Islands	Kubakota	Mary Raymond, SOAS	ELAP PhD student
	Blablanga	Radu Voica, SOAS	grant 2007-09, ELAP PhD student

**5. TRAINING FOR LANGUAGE DOCUMENTATION AT SOAS.** At SOAS we have developed several types of training for scholars wishing to undertake language documentation. This training is run by ELAP in collaboration with ELAR – some is directly funded by ELDP:

1. post-graduate courses: MA (1 year), PhD (3-4 years)

2. grantee training courses: one week
3. specialized training in recording, archiving, XML: 1 day

**5.1 POST-GRADUATE COURSES.** We offer two levels of post-graduate courses: a one year Master of Arts degree, and a three-year PhD. Students entering the PhD must have already completed an MA or equivalent.

The MA in Language Documentation and Description is a one year degree that is taught over 20 teaching weeks. The intake includes students who hold an undergraduate BA degree with a major in linguistics, along with those with no previous exposure to linguistics (generally in a proportion of 1/3 with a linguistics major and 2/3 without). The degree consists of course-work plus a short dissertation (10,000 words, essentially a long research essay), and includes a Research Training Seminar (1 hour throughout the year) and recommended attendance at fortnightly departmental seminars and occasional workshops and training courses that are held throughout the year. Starting in 2007-08 there are two pathways in the degree: a Field Linguistics pathway for students with an undergraduate major in linguistics or equivalent, and a Language Documentation and Support pathway for students with or without linguistics. The following tables summarize the structure of the two pathways.

TABLE 2. Field Linguistics Pathway

	Core	Core	Core/Option	Core
Term 1	Issues in language documentation	Technology & language documentation	Option 1	Research Foundations
Term 2	Field methods	Applied documentation & description	Option 2	Research Foundations

TABLE 3. Language Documentation and Support Pathway

	Core	Core	Core/Option	Core
Term 1	Principles of linguistic analysis	Issues in language documentation	Option 1	Research Foundations
Term 2	Principles of linguistic analysis	Applied documentation & description	Option 2	Research Foundations

The core courses are the following:

- *Principles of linguistic analysis* – an introduction to basic linguistic concepts and analytical techniques with hands-on practical exercises



- *Issues in language documentation* – this covers the topics of: project design, research ethics, IPR, researchers and communities, world language ecology, endangered and minority languages, language contact, ethnographic methods
- *Technology and language documentation* – covers information, communications and media technology basics, audio and video recording and editing techniques, data analysis and design, data formats and standards, archiving issues, software tools (Transcriber, Shoebox ~ Toolbox, ELAN, Praat, IMDI) and practical exercises
- *Applied language documentation* – introduces basic principles of applied linguistics, including orthography design, literacy, lexicography, translation, language learning and teaching, pedagogy, curriculum design, language policy, and advocacy
- *Field methods* – students work with a speaker of an unknown language to apply their knowledge and skills of data recording and analysis. In recent years the languages studied have been Khorchin Mongolian, Dida, Sylheti, and Kannada.

The Research training seminar consists of weekly meetings attended by all the students together to explore conceptual issues (e.g. what counts as explanation in linguistics, types of data, brief history of linguistic research), research methods and skills, research tools (End-Note, Powerpoint), presentation skills (including presentation dry runs), and teamwork skills. This class is especially important for cohort development and providing a forum for student concerns (e.g. workload, assessment) to be aired and addressed.

The options courses include linguistic typology, syntax, phonology, semantics, historical linguistics, language culture and society, acoustic and experimental phonetics, multimedia and language support, and areal courses (Austronesian, African languages, American Indian languages, Siberian languages). We plan to introduce a course on language revitalization in the near future as this is a topic students are keenly interested in. Students can also choose from advanced options primarily intended for PhD training (see below). These vary from year-to-year but have included grammar writing, negation, tense/aspect/mood, number, and lexical semantics.

The PhD generally takes 3 to 4 years and follows from the MA degree. SOAS offers two PhDs: a general linguistics PhD and specialist PhD in Field Linguistics. The following is the structure of this latter degree:

Year 1 – advanced level training in language documentation and description: students take 3 courses per term on advanced subjects, along with a special 3 day training course. They prepare a “core chapter” of the dissertation and an annotated bibliography, give a public seminar and are upgraded from MPhil to PhD if their work is of the required standard

Year 2 – fieldwork, normally 8-12 months

Year 3/4 – writing up dissertation, usually with a short fieldtrip (3-4 months) to check data and fill gaps

Examination of the PhD is based entirely on the dissertation and the coursework components of the degree are required but not assessed.

Occasionally, we have run special additional training events such as a three day ELAP/ELAR training program designed for PhD students and post-doctoral fellows which aim to:

- address language documentation workflow and supporting principles and skills
- develop a shared involvement in collaborative project work to enhance knowledge and skills of audio recording, processing and presentation

The following table shows the structure of the most recent such training course.

TABLE 4. ELAP/ELAR training program for PhD students and post-doctoral fellows

	Wednesday	Thursday	Friday
9.45-11am	“Managing My Data”	Audio 3: Evaluation	Images
11-11.15am	Tea break	Tea break	Tea break
11.15am-1pm	Working in teams	Audio 4: Editing	Archiving
1-2pm	Lunch	Lunch	Lunch
2-3.30pm	Audio 1	More data management	ELAN Transformation
3.30-3.45pm	Tea break	Tea break	Tea break
3.45-5.15pm	Audio 2: Practical	ELAN	Mobilization/Multimedia

**5.2 ELDP GRANTEE TRAINING. ELAP and ELAR,** in collaboration with outside specialists, offer a six-day training course for researchers who are awarded a grant by ELDP. The training course is designed to assist grantees by building on their knowledge of documentation theory and practices, especially information and media technology skills, and preparation for dealing with grant outcomes (corpus development, archiving, mobilization and publication). The following table sets out the structure of our training course run in June 2007 (the structure builds on two years of experience running these courses):

TABLE 5. ELDP grantee training program

	Wednesday	Thursday	Friday	Saturday	Monday	Tuesday
09:00 - 09:50	Welcome	Audio evaluation	Data management	Documentation	Video and documentation	Projects
9:50 - 10:40	Grantee projects				Videography	
10:40 - 11:00	Tea/Coffee					
11:00 - 11:50	Audio principles	ELDP relations	Data practical	ELAN	Video: camera	Mobilization
11:50 - 12:40		Archiving			Video: practical	
12:40 - 14:00	Lunch					
14:00 - 14:50	Digital audio	Archiving	Consultation & elicitation	Advice clinic	Video: practical	Ethics and IP
14:50 - 15:40	Audio practical	Transcription issues				
15:40 - 16:00	Tea/coffee					
16:00 - 16:50	Audio practical	Transcription practical	Field practical topics	Projects and questions	Video editing & evaluation	Wrap-up
16:50 - 17:40	Grantee projects					

Note: the shaded areas show sessions focusing on media recording, editing and management

We have run four of these training courses so far and all have had different dynamics (reflecting the experience and interest of the participants) but have all been evaluated by their participants as successful. Researchers generally feel that their existing knowledge and skills have been recognized and that they have acquired new information that they will put into practice in their own projects.

**5.3 SPECIALIZED WORKSHOPS.** From time to time ELAP and ELAR run special workshops on aspects of language documentation that draw upon the knowledge and skills of specialist instructors. We have had two such workshops to date: one dealing with audio recording, digitization and archiving, and another with extensible markup language (XML) and its role in language documentation research<sup>4</sup>. The following is an outline of the two workshops:

<sup>4</sup> XML is a document description language, used to describe the content of structured documents – each part of a structured document is described within a defined and logical structure (the structure can be documented in an XML schema or “DTD”). XML documents can be designed, created, processed and transformed manually or by using editors, stylesheets (XSLT “extensible stylesheet language for transformations”), and document processing scripts. It is the preferred data format for text materials, especially for archiving, and a number of language documentation tools, such as Transcriber and ELAN, store data in XML format.

**1. Audio Recording, Digitization and Archiving.** Presented by Professor Dietrich Schueler, Phonogrammarchiv, Austrian Academy of Sciences

TABLE 6. Audio Recording, Digitization and Archiving workshop

	Session	Topics
11 am - 1pm	Principles of audio conversion and preservation (IASA-TC 03)	analogue-digital conversion, choosing digital resolution and file formats, digital to digital ingest
	Optimizing signal extraction from magnetic tape	identifying and assessing track formats and misaligned recording heads, historical equalizations
2pm - 4pm	Field audio recording	microphone arrays, including psychoacoustic considerations, recording devices, post R-DAT
	Small scale digital archiving	manual approach to digital archiving, archiving in the field

**2. XML Day.** This workshop consisted of a mix of tutorial, practical, and discussion sessions, including group work and problem solving. It was designed to provide participants with opportunities to learn about the history, purpose, and formalism of XML, understand its applications, strengths, and weaknesses, see how XML can be applied to linguistic data, gain basic hands on experience with designing, “reading”, evaluating, and editing XML, and learn about technologies that are closely related to and used in conjunction with XML. The schedule for this workshop is given in the following table:

TABLE 7. XML workshop

10:00 - 10:40	Roots, principles and formalism. Look-ahead to architecture. Namespaces and Unicode. Workflow. Discussion of preparatory material.
10:40 - 11:20	Introduction to Oxygen. Practical exercises on document marking up and well-formedness
11:20 - 11:40	Break

11:40 - 12:40	Constraining document grammar: DTD and Schema. Constructing DTDs. XPath and XQuery.
12:40 - 13:00	More Oxygen functions. XPath exercises.
13:00 - 14:00	Lunch break (lunch not supplied)
14:00 - 14:40	XSLT and CSS, XSL:FO
14:40 - 15:20	XSLT exercises
15:20 - 15:40	Break
15:40 - 16:20	Real world: office applications and creating and manipulating XML
16:20 - 17:00	Exercises: creating and manipulating XML, bringing it all together

The content covered in the workshop included theory, practice and architecture, with the following specific topics addressed:

**Theory** XML roots, principles and formalism; XML vs. relational databases

**Practice** Case studies in representing linguistic data and theories; good XML design; XML editors; mobilization; born-XML and marriages of convenience; examples of XML in linguistics: modelling interlinear glossed text, Transcriber, ELAN files

**Architecture** Constraining XML with DTDs and Schemas; namespaces and different XML vocabularies; transforming/displaying XML with XSLT, CSS, and XSL-FO; searching XML using XPath and XQuery; Unicode and XML

**6. SUCCESSES AND CHALLENGES.** Although ELAP has only been in operation since 2003 (and ELAR since 2004) we can point to some successes in our training programs. These can be summarized as follows:

- several students who completed the MA have gone on to work in local language centers (in Australia, and Alaska), and one has recently been offered a job with an NGO in India that deals with cultural and linguistic conservation and revitalization;
- 13 students who completed the MA went on to enrol in the PhD, and all have been awarded scholarships or bursaries;

- all our PhD students have successfully applied for competitive research grants to fund their fieldwork. This fieldwork has been carried out in Iran, Mexico, Vanuatu, Senegal, Nigeria, the Solomon Islands and India;
- our current PhDs include students from Australia, India, Senegal, Sierra Leone, Sweden, Germany, Italy, and UK. Our MA students come from a wide range of countries, including US, Canada, and Australia, as well as the EU and UK;
- we have four post-doctoral fellows (two externally funded) and many visitors who together create a lively research environment for postgraduate students;
- the ELDP training courses have been well received and attain positive evaluations from attendees who have found them useful for their project work.

This success has also been accompanied by some challenges. The following are some of our current major concerns:

A lack of linguistic background for 2/3 of our MA students means that they have to work very hard to acquire basic linguistics knowledge and analytical skills at a very fast pace. We are expected to take students from a range of backgrounds, and it is usual in the UK for MA courses not to require previous knowledge of the content area. We have recently addressed this concern by expanding our Principles of Linguistic Analysis course to two terms (20 weeks) and creating two pathways in the MA degree so that the Field methods course in particular is only taken by students with sufficient linguistic knowledge.

The Field methods course is difficult to organize for the full MA cohort of 17 students to ensure that there are enough small group sessions with the language consultant. Starting in 2007-08, the number of students taking Field methods will be reduced due to the introduction of the pathways, however we may still have difficulties with this course component.

Designing and implementing courses while supervising an average of 17 MAs and 12 PhDs each year has placed a heavy workload burden on ELAP which has only two permanent staff and two post-doctoral fellows. We have received some assistance from ELAR staff who have taught course components and run specialist workshops, however the 'start-up costs' of getting the postgraduate program going have been high. The number of staff in ELAP increases to three permanent staff, one 3-year research fellow, and two post-doctoral fellows from September 2007, and this may relieve some of the work pressures being felt by ELAP staff.

We have not been as successful with PhD student recruitment as we would have liked, especially in attracting native speakers from third world countries (we currently have only two native-speaker students from Africa).

It has been difficult to obtain scholarship funding, especially for MA students. Currently SOAS charges £3,500 (approximately \$US 7,000) in fees for EU and UK students, and £10,500 (approximately \$US 21,000) for others; this level of cost can be difficult for students to pay, along with living costs in London, which are quite high.

We have had success in gaining fieldwork funding for PhD students but this is not guaranteed in the future. An additional problem is locating safe and secure field sites for students to work in. Already, one female PhD student has had difficulties with her field site in southern Nigeria, while another lived through an earthquake and tsunami in the Solomon Islands. One PhD project planned for India was cancelled due to Maoist guerrilla attacks, and we must exercise a high level of care in attending to the risks associated with fieldwork.

There is pressure from the university and the UK government for PhD students to complete their degrees within 3 to 3.5 years, however this time frame is not realistic when fieldwork and corpus development needs to be included, especially for students working in difficult field sites with previously unrecorded languages. So far, the first two of our students enrolled in the PhD are both in their fourth year and will need this amount of time to ensure submission of a dissertation of sufficient quality.

It is difficult at this point in time to fully evaluate the outcomes of our postgraduate courses and training programs. Documentary linguistics is a very new enterprise and our MA and PhD courses are unique in the world as being specialist programs designed to develop and implement theory and practice of documentary linguistics.

The future development of career paths for graduates from our MA and PhD are not yet clear. One PhD student who has almost completed has been awarded a post-doctoral fellowship in Mexico that he plans to take up in 2008, however the career paths of all our PhD students will only emerge over the next few years.

Practitioners of documentary linguistics need to improve their communication with the linguistic community and the wider world. There continues to be a level of ignorance about (and sometimes opposition to) the goals, theory and methods of language documentation among the academic linguistics community and a low level of appreciation among the general community about the value of our work. ELAP staff and students have made numerous presentations in public events to describe our work, however more needs to be done, especially to ensure continued support and funding for the documentary linguistics enterprise.

**7. CONCLUSIONS.** Documentary linguistics is a relatively new field and the practice of language documentation has really only begun in earnest in the past five or so years. We are yet to see the range of possible outcomes from this new approach, or the impact that it will have on the academic field of linguistics, on the communities of speakers of endangered languages, and on the wider world in general. Our experience so far at SOAS however indicates that the training of current researchers and a new generation of language documentation specialists will be both exciting and rewarding for all concerned, and not lacking in challenges.

## REFERENCES

- AUSTIN, PETER K. 2006. Data and language documentation. In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Fundamentals of Language Documentation*. Berlin: Mouton de Gruyter, 87-112.
- AUSTIN, PETER K. 2007. 'Survival of Languages' in Emily F. Shuckburgh (ed.) *Survival: Darwin College Lectures*. Cambridge: Cambridge University Press.
- AUSTIN, PETER K. and LENORE A. GRENOBLE. 2007. Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 12-25. London: SOAS.
- BIRD, STEPHEN and GARY SIMONS. 2003. Seven dimensions of portability for language documentation and description. *Language* 79/3:557-82.
- CSATO, EVA A. and DAVID NATHAN. 2003. Multimedia and documentation of endangered languages. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 1, 73-84. London: SOAS.
- EVANS, NICHOLAS and HANS-JURGEN SASSE. 2007. Searching for meaning in the Library of Babel: field semantics and problems of digital archiving. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 58-99. London: SOAS.
- GIPPERT, JOST, NIKOLAUS P. HIMMELMAN and ULRIKE MOSEL (eds.). 2006. *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178.)* Berlin: Mouton de Gruyter.
- GRINEVALD, COLETTE. 2003. Speakers and documentation of endangered languages. In Peter K. Austin (ed.) *Language Documentation and Description* Volume 1, 52-72. London: SOAS.
- HANS RAISING ENDANGERED LANGUAGES PROJECT. 2006. What is language documentation? Available at <http://www.hrelp.org/documentation/whatisit/>, accessed 28th May 2007.
- HIMMELMAN, NIKOLAUS P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-95.
- HIMMELMANN, NIKOLAUS P. 2002. Documentary and descriptive linguistics (full version). In Osamu Sakiyama and Fubito Endo (eds.) *Lectures on Endangered Languages*: 5, 37-83. Kyoto: Endangered Languages of the Pacific Rim.
- 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.) 2006. *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178.)*, 1-30. Berlin: Mouton de Gruyter.
- MUNRO, ROBERT. 2005. The digital skills of language documentation In Peter K. Austin (ed.) *Language Documentation and Description* Volume 3, 141-156. London: SOAS
- NATHAN, DAVID. 2006. Thick interfaces: Mobilizing language documentation with multimedia. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation (Trends in Linguistics. Studies and Monographs, 178.)*, 363-379. Berlin: Mouton de Gruyter.
- NATHAN, DAVID and PETER K. AUSTIN. 2004. Reconceiving metadata: language documentation through thick and thin. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol. 2, 179-187. London: School of Oriental and African Studies.
- NATHAN, DAVID and EVA CSATO. 2006. Multimedia: a community-oriented information and



communication technology. In Anju Saxena and Lars Borin (eds.) *Lesser-known languages of South Asia*, 257-277. Berlin: Mouton de Gruyter.

THIEBERGER, NICK and SIMON MUSGRAVE. 2007. Documentary linguistics and ethical issues. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4, 26-37. London: SOAS

WALCOTT, HARRY F. 1999 *Ethnography: a way of seeing*. Walnut Creek: Alta Mira.

WHALEN, DOUG. 2003. How the study of endangered languages will revolutionize linguistics. *International Congress of Linguists*, Prague.

WOODBURY, TONY. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language documentation and description*. Volume 1, 35-51. SOAS: University of London.

Peter K. Austin  
pa2@soas.ac.uk