ENDANGERED LANGUAGES DOCUMENTATION PROGRAMME (ELDP) APPLICATION FOR A Major Documentation Project

Read the Guidelines carefully in planning your proposal, and the Terms & Conditions of Award before completing and submitting an application.

The application form must be completed in English. ***Late or incomplete applications will not be considered.*** Submit one original hard copy with signatures which should be single-sided and unbound and submit an electronic copy. The electronic and paper copies of the application must be identical in content (except that signatures are not required in the electronic copy). Only material specifically requested in the application should be sent.

The original hard copy to be submitted by the deadline to:

ELDP Grants Coordinator
Endangered Languages Documentation Programme
School of Oriental and African Studies
10 Thornhaugh Street
London
WC1H 0XG
United Kingdom

The electronic version must be a single file in MS Word or pdf format (multiple files are not accepted) and must be emailed by the deadline to eldp@soas.ac.uk

All copies must arrive by 3rd August 2009

You should only send the information requested in the application form. If you are successful in receiving a grant you will be asked to provide the following:

- evidence that the relevant permissions and visas have been secured (if required)
- any other information required by the panel after assessment
- an assurance that an indication of support from the language community will be provided once the project has begun
- evidence of institutional pay scales used to calculate salary costs

APPLICATION FOR A Major Documentation Project Grant Ref Number: MDP

Q1 Applicant details

First Name	Jonathan	Title	Dr.
Family Name	Amith		
Primary contact address	If employer's address is to be used, provide complete details below, including email, telephone and fax numbers. Department of Anthropology Gettysburg College, Campus Box 412 Gettysburg, PA 17325 USA		
Email	jamith@gettysburg.edu		
Telephone numbers (enter information below; indicate your preferred number)			
Landline	717-337-6795		
Mobile/Cell Phone	None		
Fax	717-337-7001		

Current appointment or position

Title of position, job, or course of study	Research fellow (independent scholar)
Name of institution	Department of Anthropology Gettysburg College, Campus Box 412
Address	300 N. Washington Street Gettysburg, PA 17325 USA

How is your position currently funded?	National Science Foundation, Documenting Endangered Language Project grant: “Language Documentation Project (NLDP): Sierra Norte de Puebla”. My position is funded at about 39% full-time employment or $28,000/year. This is my sole source of revenue. The present grant, if awarded, would bring my yearly income to $52,000, without any benefits (health, pension, etc.).

Q2 Co-applicant details (complete for each co-applicant, maximum 3 permitted)

Co-applicant details

	Full name	Email	Position	Institution
Co-applicant (1)	Rey Castillo García	castagr@ hotmail.com	Asesor en capacitación de hablantes nativos (Advisor for training native speakers)	Instituto Nacional de Lenguas Indigenas
Co-applicant (2)				
Co-applicant (3)				

Q3 Host institution

The host institution which will administer the award

Name	Gettysburg College
Address	300 N. Washington Street Gettysburg, PA 17325

Principal applicant's link to the host if not currently employed or enrolled there

Gettysburg College administers my grants given that I am an independent scholar with no full-time position.

Support from your host institution – ask your sponsor to complete a statement here, maximum 500 words, demonstrating a commitment to supporting and providing facilities for your project

Dr. Amith is a research fellow at Gettysburg College and has been since his wife was hired in the Department of Anthropology in 2003. During the course of his affiliation with the College we have administered three of his grants. Two of these have been 3-year awards (totaling $600,000) from the Documenting Endangered Language program of the National Science Foundation, both for Nahuatl. The first, for Guerrero Nahuatl, ended last fall and the present, for Sierra Norte de Puebla Nahuatl, is now approaching the end of its first year. Dr. Amith has also received a small pilot project grant (£5,984) from the Endangered Language Documentation Programme ("Corpus and lexicon development: Endangered genres of discourse in Tu'un ísaví (Mixtec) of Yoloxóchitl, Guerrero"**)** that we have administered at no cost.

We have given and will continue to give Dr. Amith the facilities necessary for carrying out his present and future research. This includes an office, supplies, computer, and Internet access; full faculty library privileges and access to both interlibrary loan and our electronic databases and electronic journals; and the necessary administrative and accounting support to carry out his research and meet the terms of financial reporting for his grants. We have made a specific commitment to administer this grant at below cost, perceiving no overhead for the office space and other facilities that Dr. Amith needs to conduct his research on Mixtec, should this application be successful.

If there is any point that needs clarification, please feel free to contact me at the email given below.

Signature of representative of host institution	Name and email address John Ryan Vice-provost jryan@gettysburg.edu

Q4 Title of project

Maximum 20 words

Corpus and lexicon development: Endangered genres of discourse and domains of cultural knowledge in Tu'un ísaví (Mixtec) of Yoloxóchitl, Guerrero.

Q5	**Duration of project** State how many months of funding you are applying for (normally between 6 and 36 months)	36

Q6	**Proposed start date** (dd/mm/yy)	01/01/10

Q7	**Total amount requested** In GBP only	£133,170

Q8	**Is this a re-submission?** If Yes, provide date, reference number and type of grant originally applied for	No. However, it is a continuation of a successful Pilot Project grant.

Q9	**Other funding applications** Provide details of other funding applications or recent applications to other bodies for identical or closely related work. Include date, body, title of project, duration, amount sought and outcome if known.	For Mixtec I have no grant application pending. I may apply to the NSF to meet the 15 September 2009 deadline for the Documenting Endangered Languages program. For language documentation of Nahuatl I applied (Sept. 2007) to the National Science Foundation, Documenting Endangered Language program and was awarded a three-year grant in May 2008. The project began in August of that year. I receive $28,000/year from this grant (estimated at 39% FTE).

Q10 Commitment to the project

The intended average number of hours and overall % time per week to be devoted to the project

	Hours	% Time per week
Principal Applicant	13	33%
Co-applicant (1)	40	100%
Co-applicant (2)		
Co-applicant (3)		

Q11 Summary of proposed project

This information will be used on our website if you are funded. Include details of the language community, the country and region, and an estimate of number of speakers. Give the ISO-639 code and geographical reference e.g. latitude/longitude or Google map reference of the language community.

This is a summary (maximum 100 words); full details of your proposal should be given below

This project focuses on endangered genres of discourse and threatened domains of cultural knowledge in Yoloxóchitl (16º 48' 58"N, 98º 41' 12"W) Mixtec (*xty*), spoken in four villages within a 12 km radius in coastal Guerrero, Mexico. It will build upon a successful pilot initiative (100 recordings and time-coded transcriptions totaling 38 hours; a 1,500-stem lexicon) to produce an additional 110 hours of ethnographically rich recordings and parsed, glossed, and freely translated transcriptions. The lexicon will minimally comprise all lemmas in the corpus. The results will create the first extensive, archival quality Mixtec corpus and establish a foundation for future Mixtec studies, particularly in phonetics, phonology, and syntax, areas in which Yoloxóchitl Mixtec (YM) is of typological interest.

Q12 Specific outcomes (Maximum 750 words)

The multidisciplinary project will employ the methodology that Amith, the project director, has utilized in his previous projects on Nahuatl, in which he has focused on (1) endangered genres of discourse and (2) threatened domains of cultural knowledge, particularly rapidly disappearing expertise on the natural environment, material culture, and past traditions. Amith will also apply a methodology of *cultural lexicography*, which stresses linking those lexical entries best described in an encyclopedic manner to exegetical texts (recordings and transcriptions) about the targeted lemma. Using this methodology Amith has recorded over 2,000 Nahuatl texts of varying lengths (totaling over 150 hours) on local flora and fauna, the heavens, disease and cures, food preparation, agricultural and hunting or fishing practices, and children's games, among many other themes (see Amith CV for sample materials posted at http://www.balsas-nahuatl.org/documentation). Documentation of discourse on quotidian activities assures both ethnographic richness and the inclusion of narrators of both sexes and all age groups. The methodology that Amith has utilized in his documentation efforts for Nahuatl has led to a corpus of materials that Heidi Johnson (appendix, p.2), AILLA project manager, has called "possibly the most diverse [set of materials] in the archive, apart from AILLA Director Joel Sherzer's Kuna Collection". This methodology will produce the same richness of material for YM—different genera and topics recorded by individuals differing in age, social status, and sex.

Indigenous knowledge of local flora and fauna is a particularly endangered realm, yet one that is seldom part of language documentation efforts: ethnobiologists rarely know the local language and language documentation specialists seldom have the skills or professional contacts to collect and identify specimens. This project is unusual in that the assembled team possesses the skills to link ethnobiology and documentation. After a decade of documentation and ethnobiological research, Amith has built up a collaborative network of over 100 botanical and zoological experts (listed in appendix, pp. 10–13), who have provided scientific determinations of over 2,250 plant and animal specimens collected in Nahuatl-speaking areas. This network, along with institutional support at the Smithsonian Institution's botany and entomology departments and the Universidad Nacional Autónoma de México (appendix, pp. 7–9) ensures that specimens will be professionally collected, identified, and curated. By linking native language exegetical texts and scientific determinations in Western nomenclature, the linguistic and cultural documentation of flora and fauna becomes immediately accessible and relevant to linguists, anthropologists, and (ethno)biologists. Moreover, a focus on this endangered domain of knowledge is of great interest to the YM community (see support letter, appendix, p. 1). Professional quality photos and illustrations (see appendix, p. 17) will be made available to this community along with the relevant recordings and their transcriptions.

Effective documentation requires not only best practice procedures, from quality recording, to accurate time-coded transcriptions and a detailed metadata registry (e.g., based on OLAC standards). It requires the phonological, morphological, and semantic *transparency* of the transcriptions that would facilitate future research. This project will provide this transparency. First, all orthographic conventions will be clearly articulated so future researchers will understand the graphic representations used. Second, the transcriptions will be parsed and glossed to facilitate dictionary lookup. Much of the parsing will be carried out in the transcription stage: following standard conventions for Mixtecan languages, clitics will be separated from stems by the equal sign (=) and affixes by a hyphen (-). Additional stem-internal parsing (e.g., of the contour tone that may express the negative) will be accomplished through a well-known parsing program (the Xerox Finite-State Transducer), which will be built for YM morphology by William Poser, working in conjunction with Amith and Castillo (see research staff, p. 23). Combined, the orthographic conventions for expressing clitics and affixes and the implementation of a transducer built for YM will enable the processing of all transcriptions to produce an archival four-line interlinear text format that will facilitate future linguistic and anthropological research.

Yet even a parsed and glossed time-coded transcription may be opaque to future researchers. To avoid this, the present documentation project will archive:

1) a clearly stated ontology for all glosses;
2) a dictionary (see sample entry, appendix, p. 18) that contains at a minimum all words in the corpus along with detailed semantic descriptions that include set phrases and collocates,

particularly important in a relatively isolating language such as YM, in which many semantic concepts are expressed by short set phrases or collocations;

3) a brief grammatical sketch to facilitate more in-depth research by future scholars.

In sum, the present project will produce high quality documentation by providing:

1) An ethnographically rich corpus comprising 110 hours of time-coded transcriptions in four-line interlinear format (surface, parse, gloss, and free translation into Spanish) and 140–190 hours of additional recordings not selected for transcription.

2) Varied genres of discourse and domains of knowledge.

3) Diversity in the sex, age, social status and origin of individuals recorded.

4) High fidelity recordings with good headworn microphones and best practice specifications: 48KHz, 16-bit.

5) Full and accurate metadata.

6) Time-coded transcriptions in four-line interlinear format.

7) A description of all conventions (e.g., orthographic) and terminology (e.g., glosses) used.

8) An extensive, semantically detailed bilingual dictionary (YM/Spanish), minimally comprising all lemmas in the corpus and with great attention given to set phrases and idioms.

9) A basic reference grammar to orient future researchers

10) Access to materials for both scholars and members of the native community through permanent archiving and local availability in a combination of printed and CD formats (for community involvement, see section on community contexts).

11) Online access to all materials: dictionary, grammar, and texts.

Q13 Detailed description of the project

(Maximum of 3,000 words across the 5 headings)

Language context

Introduction
This project will *build upon* a successful ELDP pilot project to create the first extensive, archival quality corpus of recorded and expertly transcribed material in a Mixtecan (Tu̲'un ísa̲ví) language.

State of the language
Mixtec is here considered to be a language *family*, part of a larger unit, Otomanguean, that Suárez (1983:26) considers to be "a 'hyper-family' or 'stock'." Mixtecan languages (spoken in the states of Oaxaca [156 municipalities], Guerrero [13 municipalities], and Puebla [10 municipalities]) are highly varied internally, the result of approximately 2,000 years of diversification. Estimates of the number of Mixtecan languages vary (*Ethnologue* lists 53, Smith Stark [1995] mentions 45; Suárez estimates about 29 distinct languages; Bradley and Hollenbach [1988:1] suggest "perhaps twenty unintelligible languages") as do the criteria utilized for such determinations (mutual intelligibility is favored by SIL and *Ethnologue*; others [e.g., Josserand, 1982] use lexicon, morphology, or isoglosses). Thus, the documentation of one or several Mixtecan languages would not preclude the need to document others but, as the next section demonstrates, adequate documentation of even one Mixtecan language has not been produced.

Among Mixtecan languages, those spoken in Guerrero are greatly underrepresented in existing phonological and syntactic descriptions (see following section). Castillo's unpublished masters thesis is the only description of Yoloxóchitl Mixtec (YM), with approximately 3,000 speakers in Yoloxóchitl and another 1,500 in Arroyo Cumiapa (about 6.5 miles to the northeast), both towns with a fairly high degree of vitality. YM is also spoken, though rapidly disappearing, in Cuanacaxtitlan (pop. about 4,000) and Buenavista and its surrounding rancherías (total pop. about 5,000), though in both areas language loss and very high relexification is virtually complete among the younger generation. It is fair to say that in Cuanacaxtitlan and Buenavista the "tipping point" has been reached and the language has passed from one side to the other of what Fishman (1991) calls "the continental divide," the point at which intergenerational transmission of language ceases and maintenance becomes an often fruitless endeavor. In the other two communities—Yoloxóchitl and Arroyo Cumiapa—the situation is slightly better. Nevertheless, YM is highly endangered and the current eldest generation is probably the last with expertise in the range of endangered genres of discourse and threatened domains of cultural knowledge that is targeted in the present project.

YM is further threatened by unstable bilingualism, close proximity to a Spanish-speaking municipal capital (San Luis Acatlán), and emigration to national and international urban destinations. Relexification, the loss of terminological repertoires for specialized activities, and the continual disappearance of ritualized texts of various genres exacerbate loss. These factors give urgency to a documentary effort, particularly if the lexical and syntactic richness of the language is to be recorded.

Extant material on Mixtecan languages (see list of archival holdings and major publications, pp. 54–58)
There are almost no narrative, transcribed recordings in a Mixtecan language (see bibliography) beside the 38 hours (100 distinct items; available at http://www.balsas-nahuatl.org/soas see folders under each narrator) of high quality digital recordings and time-coded transcriptions produced by Amith's and Castillo's pilot project with minimal resources (£5,984 from ELDP and $10,000 from the Ford Foundation, used mostly to support Castillo). The two major archival holdings are Josserand's archive at AILLA and Macaulay's and Hinton's materials at Audio Archive of Linguistic Fieldwork, Berkeley Language Center. In regard to the former, of the 318 individual items archived, 164 are elicitation and 139 are wordlists; all but two items are from Oaxaca. Of the six narratives with a total time of 27 minutes 43 seconds, at most half of the time is Mixtec (the remaining is Spanish translation or discussion). In regard to the Berkeley holdings, there are two sets of materials, both from San Miguel el Grande, Oaxaca. Most items are listed as elicitations or miscellaneous words and phrases. There are very few narratives and apparently no transcriptions.

There is also a dearth of printed Mixtecan language texts. Dyk (1959) provides one set of materials (approximately 20,000 words) and Hollenbach (1988) another (approximately 3,500 words; see a few additional items listed in the bibliography). In neither case is a sound recording of the textual material available.

In sum, a review of academic publications and archival holdings on Mixtecan languages reveals a clear lack of primary documentation materials (sound recordings and accompanying transcriptions), a situation that becomes even more disquieting given that the primary foci of academic interest in this family of relatively isolating VSO languages has been on (1) phonetics and phonology, and (2) syntax. Both these areas of research would benefit immensely from the material that this project will provide: a large and diversified corpus of high quality digital recordings, accurate time-coded transcriptions in four-line interlinear format, a semantically rich dictionary comprising all lemmas in the corpus with correct presentation of lexical tone, and a grammatical sketch.

Typological interest of the language

Mixtecan languages are particularly interesting for their phonology (especially tone) and syntax. In a recent article, John Daly and Larry Hyman (2007:165) note that "the complexity of Mixtec tone systems has been recognized for some time." Indeed, many works (see bibliography) have explored phonology, most often tone and vowel features (nasalization and glottalization). YM is striking for its tonal inventory. It has (Castillo 2007) five tone levels and 19 different tonal patterns on bimoraic tone-bearing units (7 of which have contour tones on one mora. It is also apparently unique in its innovative use of tone alone to mark the 2nd-person singular on verbs with a high or mid tone on the final mora). Note that Chalcatongo Mixtec has only 3 levels and 9 tonal patterns in the bimoraic couplets, though 5 patterns account for 93% of the lexicon. If Castillo's analysis can be documented and the material (recordings and transcriptions by a proficient native speaker) made available, it could significantly impact our general understanding of Mixtec tone.

Equally interesting from a typological perspective is the basic VSO word order of Mixtecan languages, an order much less common than verb-final (Dryer, n.d.:3). Syntax is explored in Bradley and Hollenbach's four-volume compilation though the focus is not typological. Macaulay (2005) demonstrates the typological relevance of Chalcatongo Mixtec as a verb-initial language and explores the manner in which this language manifests many of the expected typological correlates of basic VSO word order. The present project will continue to explore these questions and significantly expand the primary data available for research on VSO languages and typology.

Despite the research on tone and syntax in Mixtecan languages, there is little primary material available to researchers and apparently none that would rival even the 38 hours of recordings and transcriptions that Castillo and Amith have already produced. Research in both phonology and syntax will be greatly advanced as this project's results are made available: a large corpus of actual speech and a time-coded transcription with complete and accurate tonal marking, an accompanying detailed lexicon and, eventually, interlinear representation of parses and glosses (including part of speech tagging). The section on research outcome further discusses the importance of the present project for linguistic research.

Cultural facets of Yoloxóchitl Mixtec documentation

For all threatened languages, disappearance is a complex affair. Some languages die with, to quote Nancy Dorian, "their boots on", the last speaker still fluent, though out of practice. Other languages undergo slow linguistic change that may finally involve extensive relexification and morphosyntactic simplification. Yet besides the loss of linguistic forms (lexemes, morphemes, and syntactic complexity) there are significant cultural concomitants to language death, what Campbell and Muntzell (1989:195) have called "stylistic shrinkage" or "functional deprivation". That is, language loss proceeds at an uneven pace in different domains of use.

This project, the director of which is an anthropologist, has been specifically designed to target endangered genres of discourse and threatened domains of cultural knowledge. This is reflected in methodological practice (see section Q12), the differing expertise of the two groups of language consultants that will be asked to collaborate (see section Q16d), the vast network of botanical and zoological advisors and colleagues with whom Amith has already worked and who will continue to support the ethnobiological facets of the project, and, most significantly, the interdisciplinary composition of the research team.

Experience and qualifications of the team

The interdisciplinary research team that has been assembled comprises experts in ethnography, language documentation and lexicography (Amith), Mixtec morphosyntax and phonology (Monica Macaulay),

phonetics of tonal languages (Heriberto Avelino), computational linguistics and programming as well as phonetics (William Poser), and ethnography (Amith). Most importantly, the team includes Rey Castillo, a native speaker of Yoloxóchitl Mixtec with a masters degree on YM phonetics and phonology, who has been working with Amith on YM documentation for the past two years. In each case, the individuals chosen are among the most qualified and experienced scholars to carry out the present project.

Amith has been working on indigenous language documentation for more than a decade. He has been documenting two variants of Nahuatl (Balsas Valley, state of Guerrero, and the Sierra Norte de Puebla, state of Puebla) and, during the past two years, Yoloxóchitl Mixtec. He is thoroughly familiar with best practice principles, from fieldwork to archiving. Moreover, his training as an anthropologist has motivated a highly ethnographic methodology in his documentation efforts. Macaulay is one of the worlds leading experts on a Mixtecan language (Chalcatongo) and has written on phonology, morphosyntax, and typology. Avelino is one of the few phoneticians who has worked on an Otomanguean language. Poser is a renowned phonetician and computational linguist (and a member of the technology advisory committee to the Linguistic Society of America) who has already contributed several programs (Fix Mixtec, ShoePolish, Prompter/Segmenter; see his CV) to this project. Castillo is one of the few individuals capable of an accurate transcription of a Mixtecan language, with accurate tonal markings. He has worked with Amith over the past two years on a YM documentation effort. Together they offer complementary skills that has been key to producing a unique set of materials for a Mixtecan language. After the completion of a major documentation project, should it be approved, Castillo will pursue doctoral studies in linguistics with a focus on Mixtecan languages and use the material produced by this project in his doctoral research.

Documentation methods

Recording standards and metadata

All recordings will be digital at a 48KHz sampling rate and 16-bit word size (24-bit may be used for music) utilizing an external headworn cardiod microphone with XLR connections. For the pilot project the recorders were a Sonifex Courier and Marantz PMD 670 with Audio Technica ATM 75 condenser headworn microphones. Recording is mono for a single speaker, stereo for two-person exchanges. All files are stored in three locations—two external hard drives and a commercial server rented by Amith—before final transfer to the ELP archives. Copies are always maintained on the original three locations.

Metadata is recorded in Toolbox and exported to XML for permanent archiving (the metadata for the pilot project—100 items and 38 hours—has been sent to David Nathan at SOAS along with the digital audio files).

For informed consent of narrators, see section Q14.

Fieldwork practice

The focus on endangered genres of discourse and threatened domains of cultural knowledge and the methodology of cultural lexicography were described in Q12. The strategy of working with two sets of language consultants ([1] those who provide narrations, stories, and exegetical texts on quotidian activities and [2] experts on the natural environment) is described in section Q16d. Approximately 30–40 days/year will be dedicated to recording, yielding between 75 and 120 hours of material per year.

In addition, elicitation and interviews will be used to target specific issues (such as tone sandhi) in Mixtecan languages that are of theoretical interest (see the section on research outcomes below).

Fieldwork will be carried out by Castillo (approx. 6 weeks/year), Amith (approx. 4 weeks/year), and Avelino (approx. 2 weeks/year). A team of biologists from Nelly Diego's lab at the Universidad Nacional Autónoma de México (appendix, p. 7) will spend approximately 3-4 weeks in the field each year.

Transcriptions, Lexicon, Grammar

The pilot project revealed that Rey Castillo can transcribe approximately 1.5 hours of recordings/week (including revision and proofing). He will dedicate 50 percent time to transcriptions, yielding about 110 hours over the course of the three-year project. Work on the lexicon in the pilot project also demonstrated that together Amith and Castillo can elaborate 15 entries/day (with Castillo working full-time and Amith consulting and reviewing all entries). A 2000-stem dictionary (with many entries having extensive subentries) would require 135 days (27 weeks or 18% of project time). Castillo and Amith

will work 1 day/week in writing a basic reference grammar (this represents 50% of Amith's time on the project and 20% of Castillo's). Castillo will spend the remaining 12% of his time (about 6 weeks/year) in the field with the different researchers—Amith, Avelino, and the collaborators in ethnobiology—with much of this time dedicated to recording.

Time-coded transcriptions will be elaborated in Transcriber or ELAN using a practical orthography and Unicode encoding. To facilitate typing, no keyboard mapping is used. Rather, the final text will be converted to a more visually friendly script using a program, Fix Mixtec, developed by William Poser (a member of the current team). At present this converts the practical orthography used by Castillo to an IPA representation, though any target graphic system may be established. Finally, practical language materials will be formatted in paragraph for local distribution and potential publication.

The lexicon (bilingual YM/Spanish) is being built in Toolbox and will be exported to XML. William Poser will write a script for converting this XML database into a dictionary format (e.g., through XSLT). The integrity of the database (e.g., that all entries have a valid and unique numerical identifier [UID], that no entries are repeated, that all cross-references are valid) will be tested by Poser's ShoePolish program.

Mixtecan languages are relatively isolating and thus many semantic concepts are expressed through set phrases or collocations. Thus the significant statistic on lexicosemantic thoroughness is not the number of entries (almost always bimoraic stems) but rather the number of set phrases and collections (usually two, at most three, words) with distinct semantics. In the YM dictionary these phrases and collocations are structured as subentries. This is illustrated by a sample entry (from the incipient dictionary being created by Castillo and Amith) for *i³ni²* (heart), which contains 24 senses, all but two as set phrases (see appendix, p. 18). For example, note the ninth subentry, *ka³ka³ i³ni²* | to doubt (literally 'to walk around' + 'heart'). In such cases the lexicographer must decide where to enter the item: *ka³ka³ i³ni²* is entered under *i³ni²*. It is important, however, that the phrase be discoverable through either the collocator or base. The database allows this by using different XML tags for each word of the phrase. The element that is the same as the headword is tagged with <mix/>, simply indicating a YM word. The other element is tagged with <vmix/>, which has multiple functions: (1) in an online dictionary it both allows the value (in this case *ka³ka³*) to be used as a hyperlink to the full entry under *ka³ka³* and it facilitates queries that discover words as either headwords or collocates; (2) in a printed dictionary the separate tagging facilitates the generation of an extensive index or cross referencing system (e.g., under the headword *ka³ka³* a list of all phrases in which this word appears could be easily generated).

Archiving

Archived recordings are always accompanied by extensive metadata in an XML document. The quality and diversity of Amith's archived material is mentioned in a support document from Heidi Johnson included in the appendix (p. 2). David Nathan of ELP has also seen the results of Amith's pilot project with Rey Castillo on YM. The material can be accessed and downloaded at http://www.balsas-nahuatl.org/soas and then by narrator (subfolders exist for sound files and time-coded transcriptions).

Community contexts

In their pilot project, Amith and Castillo demonstrated a commitment to collaborating with YM communities and to making the documentation results locally accessible (see receipt of materials by Yoloxóchitl authorities, appendix, pp. 14–15). As a result, the community has expressed renewed support for a documentation project and confirmed its disposition "to collaborate to the extent its means permit to create a collection of linguistic and cultural data including that on the flora and fauna of the region, all of which will be placed at the disposition interested parties in the public library of our community" (see letter, appendix, p. 1).

An important facet of any documentation project is collaboration with native speakers who can most effectively carry on work after the project ends. This project will advance the training and academic career of Rey Castillo, a native speaker of Yoloxóchitl Mixtec with a masters degree in linguistics and two years of part-time experience working with Amith on YM documentation. Castillo's work with a team of individuals, each with an area of expertise (Amith on documentation techniques and lexicography, Avelino and Poser on

phonetics and phonology, Poser on computational linguistics, Macaulay on Mixtecan morphosyntax), will enhance his preparation for doctoral work in linguistics and enable him to perfect his research skills and develop substantive documentary materials that he will use in his doctoral studies in linguistics. Additionally, Castillo, a teacher by training with expertise in indigenous education, will also conduct periodic workshops in Yoloxóchitl on the results of this project in an effort to stimulate local interest in literacy and revitalization. He will consult with community authorities and educators to determine the best format for practical language materials that will be produced by this project (supported by a budget of £2,000).

In his NSF- and Ford Foundation-funded documentation initiatives on Balsas Nahuatl (state of Guerrero) and Sierra Norte de Puebla Nahuatl, Amith has demonstrated a commitment to working collaboratively with indigenous villages, transferring documentation skills to native speakers, and striving to make the products of documentation efforts available to the communities (see support letter from Fernando Nava, director of the Instituto Nacional de Lenguas Indígenas, appendix, pp. 5–6). For example, Amith has produced a set of 6 CDs (over 6 hours) of recordings from three villages in the Balsas valley and prepared transcriptions for publication in a book of over 200 pages. INALI has published 10,000 sets (book and CDs) that will be distributed freely to Nahuatl-speaking communities in central Guerrero. INALI's director has expressed interest in publishing similar material in Yoloxóchitl Mixtec.

In an ongoing three-year project, Amith is collaborating with Tosepan Titataniske (see support letter, appendix, pp. 3–4), an indigenous collective from the Sierra Norte de Puebla. He is training three native speakers, who work full-time. The team, which also includes a botanist, is working with Tosepan to establish a local museum, herbarium, and cultural center, to be administered by Tosepan. Here the results of the language and cultural documentation project will be made available to the local community and to outside students and scholars.

In sum, the efforts of Amith and Castillo during the SOAS pilot project demonstrate a level of collaboration with the Yoloxóchitl community that will continue and be enhanced in a major documentation project, should it be approved. Amith's track record in training native speakers, making materials available locally, and collaborating with native communities is confirmed in the support letters from INALI and Tosepan (appendix, pp. 5–6 and pp. 3–4) and will be a key methodological and ethical component of work in Yoloxóchitl Mixtec.

Budget

Detailed budgetary expenses are listed and justified in the subsections of Q16. Equipment costs are virtually nil, with most equipment having been acquired for previous documentation projects or, for Avelino, available through the Max Planck Institute. The administrative costs of the project, £4,586, are extremely low (under 3.5% of the total budget of £133,170). The PI and all senior US researchers (Amith, Poser, and Macaulay) are budgeted at £22/hour, a modest rate of compensation considering the lack of any fringe benefits (health insurance, taxes and benefits). The Mexican collaborators, Castillo and Mendoza, will be paid according to the base scale utilized at the Universidad Nacional Autónoma de México for individuals with the same skill levels.

Research outcomes

This project will document a virtually unstudied Mixtecan language—Yoloxóchitl Mixtec—by creating a complete and extensive set of primary materials (see section Q12). Methodologically, this project offers several innovative approaches, particularly the use of a parsing program to help generate three-line interlinear time-coded transcriptions (the free translation will be added), the use of XML tags in the dictionary to facilitate cross-referencing of set phrases and collocations in this relatively isolating language, and fieldwork practice ("cultural lexicography") that is heavily ethnographic and stresses targeting endangered genres of discourse and threatened domains of cultural knowledge (here the support team of biologists is of world-class specialists). It will also develop, through the participation of William Poser, software specifically designed to advance documentation practice (see Poser CV). Yet the project will go beyond simple production of primary materials (corpus and lexicon) and the development of an innovative methodology. It will address important

descriptive and theoretical issues in the areas of phonetics and phonology, and morphosyntax, that have been central to linguistic research in Mixtecan languages. It will best accomplish this through collaborative work among a team of linguists, anthropologists, native speakers and programmers (Amith, Avelino, Castillo, Poser, and Macaulay).

Phonetics and phonology

Two major topics in phonetic and phonological studies on Mixtec are tone and nasalization. Another important though understudied phenomenon is the acoustic nature of stress and its relationship to tone. The present project will advance understanding of these central issues by analyzing extensive primary documentation (recordings and time-coded transcriptions) and by carrying out fieldwork focused on the targeted research topics.

Tone: The complexity of the tonal structure of Mixtecan languages is of significant interest (see bibliography; Daly and Hyman, 2007:165). Some of the most important studies have focused on floating tones (dating from Pike's 1944 study), sandhi (e.g., Hunter and Pike, 1969), and the relative frequency of different tonal sequences in the bimoraic tone-bearing unit (e.g., Hinton 1991; Macaulay 1996). Floating tones are word-final high tones that are overtly manifested only when the word so marked is followed by another with certain tonal patterns (for Chalcatongo, this includes the absence of high tone in the first mora of the following word; see Macaulay 1996:25ff). This project will inventory all YM words that (given comparative data) might be expected to manifest floating tones and it will study the impact of these words on the pitch contour of that which follows. This study will also explore the occurrence of sandhi beyond that described in Castillo's thesis, the only source of information on YM. Castillo found only a few cases of sandhi. One such case affects a set of three enclitics (*=ra*^{3}, 3sg; *=ndu*^{3}, 1pl.excl; *=na*^{3}; 3pl), the tones of which are lowered (*=ra*^{1}; *=ndu*^{1}; *=na*^{1}) when following a stem that ends in tone 2 or 1 (but not 5 or 3; we use a superscript 5 for high tone, a superscript 1 for low tone). The tone of the same enclitics also changes in different phrasal positions: when not phrase-final, the 3 is reduced to 2 (*=ra*^{2}; *=ndu*^{2}; *=na*^{2}).

Finally, studies of Mixtecan languages have looked at the relative frequency of diverse tonal sequences over the tone-bearing unit and have demonstrated the uneven distribution of different patterns. The present project will carefully determine the lexical tone of all words. The initial determination will be established by Rey Castillo, a fluent native speaker who for his masters degree in linguistics wrote a thesis on YM phonology. Avelino, Castillo, and Poser will then work together to develop a methodology for eliciting and establishing lexical tone. They will carefully determine whether elicitation frames should be used and, if so, what they should be. Poser's program *Prompter/Segmenter*, will be used to elicit spoken word tokens from the lexical database, insert them in a frame if deemed necessary, and then help segment the target word recording into individual token files named with the unique ID of each word. Avelino and Castillo will be responsible for the acoustic analysis. The result will be a highly accurate (combining native speaker knowledge with phonetic analysis) representation of lexical tone and a lexical base for studying floating tones and sandhi. Poser will process information to generate statistics on the frequencies of each tonal pattern. The project will archive the segmented word-token sound files, the lexical database, and the raw data of acoustic analysis. Future researchers will thus have available all the primary and secondary material to reevaluate any of our conclusions. Finally, YM has been analyzed by Castillo to have five tone levels and a 19 different tonal patterns on bimoraic tone-bearing units. If this assertion can be documented and the material (recordings along with transcriptions by a proficient native speaker) made available for research, it could significantly impact our general understanding of Mixtecan tone.

Nasalization: Nasalization appears to be more limited in YM than in many other Mixtecan languages in which this phenomenon has been studied. For example, in Chalcatongo (Macaulay, 1996) all vowels share the feature for nasalization regardless of the intervening consonant; in Ñumí Mixtec nasalization affects all segments (vowels and consonants), a pattern that motivated Martlett (1992) to consider it an autosegmental phenomenon. Coatzospan is the Mixtecan language with the most complete phonetic description (Gerfen, 1996). Nasal vowels, which are limited to word-final position, undergo a process of regressive assimilation, spreading to the first mora in CVV and CV'V stems and in bimoraic stems when the intervocalic consonant is voiceless. Additionally, vowels following, but not preceding, a nasal are nasalized. Avelino, Castillo, and Poser will develop a research framework to study YM nasalization; Avelino and Castillo will carry out fieldwork (Avelino, a research at the Max Planck Institute, has access to the necessary equipment) and all three researchers will prepare a preliminary analysis for publication, thus contributing to the widespread literature on this topic.

Stress: The material generated by this project (primary documentation and targeted elicitation) will be used to analyze the nature of stress in YM. Castillo has not noted stress in his phonological study of YM, so the question of its salience is a topic for research. Gerfen (1996) presents evidence that amplitude and length are the phonetic cues to stress in Coatzospan Mixtec and that in phrasal constructions stress is only found on the final tone-bearing unit (this is not the case in Chalcatongo; Macaulay 1996:110). Avelino, Castillo and Poser will collaborate on a preliminary analysis of this phenomenon although this project does not aspire to present a complete study of stress in the tonal YM language. But it will produce and archive material, again under the guidance of Avelino, Castillo, and Poser, that will allow future researchers to investigate stress.

In sum, the present project will explore topics in the phonetics and phonology of YM that have received widespread attention in Mixtecan studies and that are of general theoretical interest. Castillo, Avelino, and Poser will work to ensure that the representation of lexical tone is accurate (and can be checked by others in the future against recorded data). The project will present a statistical analysis of the distribution of tone across the tone-bearing unit, it will explore sandhi and floating tones. It will also carry out a preliminary study of, nasalization and stress and create primary archived materials, thus facilitating future research (e.g., by referencing any analysis to the primary data on which it was based).

Morphosyntax

Research into the morphosyntax of any language is greatly enhanced by the availability of a large lexical database along with a recorded corpus and interlinear, time-coded transcriptions. The present project, by creating such a material (along with an extensive, detailed dictionary), will provide a significant means to advance our understanding of the morphosyntax of a Mixtecan language. Our research goals target three topics that have been particularly challenging: 1) the relationship between the two basic verbal aspectual stems; 2) the semantics of the verbal aspect system; and 3) non-basic word order.

YM verb stems manifest two basic aspectual forms: the potential and the habitual (or realis). In YM the formal relationship between the two major aspectual stems is varied, but *always present* (and in this way different from Chalcatongo Mixtec). In YM at times the potential and habitual differ simply in tone (*ka*^{3}*ku*^{3} vs *ka*^{5}*ku*^{3}; 'to be born') and at other times in both segments and tone (*ka*^{3}*ka*^{3} vs *xi*^{5}*ka*^{3}; 'to walk around'). The patterns of relationship between the two aspectual stems of a verb are probably not arbitrary; indeed, they suggest a still poorly understood historical derivational process (e.g., a single segment prefix on a protolanguage VCV stem). Yet the lack of a large, easily analyzed dataset has inhibited study of this relationship. The present project will provide the needed extensive lexical database with both verbal aspectual stems clearly presented and easily compared. Poser will write a simple program to statistically summarize the phonological relationship between the two stems for all the verbs in the lexicon.

Analyses of the use and meaning of the two major aspects—habitual/realis and potential— are few and limited (see the sketches in Hollenbach and Bradley, eds., 1988–92; Macaulay, 1996:chap. 3, *passim*). None are based on an extensive corpus of natural speech. This project, by producing a corpus of interlinear transcriptions with verb stems parsed out, will greatly facilitate an analysis of the range of use and meaning of the two basic verbal aspects.

Finally, YM is a VSO language, manifesting a relatively uncommon basic word order of significant typological interest. Macauly (2005), however, is virtually the only author who has presented material on a Mixtecan language (that spoken in Chalcatongo) to test the universality of the correlates of VSO basic word order proposed by Greenberg and others. The present project will explore similar issues of typological concern with data from YM. The project will also produce a large corpus that will permit a reexamination of Macaulay's (2005) assertion (again, for Chalcatong Mixtec) that initial subjects and initial obliques are fairly common while initial objects are fairly rare.

Non-basic word order occurs in both topic and focus constructions, syntactic structures that would be expect to occur repeatedly in a large corpus. Much more study is needed on these constructions in YM. The present project will provide the means (exegetical texts and targeted fieldwork elicitation when necessary) to address this research topic, including an analysis of any prosodic implications of topic and focus constructions. As with all other research, the project will archive both the research results and the primary material on which these results were based.

In sum, syntactic studies of Mixtec will be greatly enhanced by a large corpus of natural language materials in digital audio and time-coded text format with an accompanying lexicon and, eventually, interlinear representation of parses and glosses (including part of speech tagging) and a free translation.

Q14 Ethical aspects of the project

Outline how you intend to handle ethical issues (Do no exceed this page of A4 paper)

Ethical issues arise at three major points during the documentation process: (1) recording and transcribing; (2) archiving; (3) dissemination. These issues will be addressed following the approach that Amith has successfully applied in previous projects.

Recording and transcribing: Before recording, the academic, archival, and educational goals of the documentation project are carefully explained to all narrators in their native language. They are told that the recordings and, in most cases, transcriptions, will be archived in a foreign institution (in this case the ELDP and AILLA archives) and will be heard by scholars. They are also told that the material may be used by students and second language learners. They are asked if they would also permit local copies and copies in other communities or schools where their mother tongue is spoken. A separate section of the agreement asks permission for local community access. The agreement also ensures that absolutely no commercialization of the recording will be allowed and that anyone who wants to use the material must sign a separate agreement with the narrator. All narrators are given a written "contract" (in Spanish) explaining the project and their intellectual property rights. The agreement is always communicated in the native language as well. The contract is signed by a documentation project member (usually the one doing the recording) and by the narrator (if unable to sign a family member is asked to be a witness). Recording is not carried out without a signed (or in a few cases oral) agreement. The narrator is given a copy of the contract and a CD of the recording, often burned on the spot. In some cases the disk is given later.

Transcriptions are true to the recording. However, if in transcribing the project members note sensitive personal statements (which may occur in life history narratives) any agreement for local archiving and distribution is revisited with the narrator. At times it is possible to edit out (in the recording and transcription) the potentially embarrassing sections.

An example of the agreement (in Spanish) used in Mixtec documentation projects is included in the appendix (p. 16).

Archiving: Permanent archiving outside the communities must respect the terms requested by narrators who, in over a hundred cases up to the present, have not requested any restrictions on access. Local archiving requires a special and clear statement by the narrators that their material may be kept locally. In the case of the Mixtec pilot project all narrators have so agreed.

A document signed by the village authorities (one of whom, Maximiliano Francisco, was a narrator as well) certifies the receipt by the Yoloxóchitl community of 10 CDs with 44 recordings by 5 speakers. Additional material will be given to the community at the end of summer 2009.

Dissemination: Particular care is given to transparency in the dissemination of recorded and transcribed material. That is, even though at the time of recording a narrator may have agreed to the utilization of his or her material in local schools or in a freely distributed government publication (as has occurred with Nahuatl materials), at the time of local deposit or publication, the narrator is approached a second time to ensure that he or she is fully aware that the recording and transcription will be made available, be it in local schools, distributed in dozens of indigenous communities, or published for educational use.

Community support: Amith and Castillo have been working with the Yoloxóchitl community for close to two years. About half the recordings have been given to the community (with the narrators consent). The community authorities have expressed strong interest and support for continuing documentation efforts as demonstrated by a letter of support extended to Amith and Castillo on 15 May 2009 in which they offer, within the realm of their possibilities, to collaborate in the project.

Q15 Project work plan

Provide a work plan (do not exceed this page of A4 paper) to show the steps and timeline of the work that you will undertake in the project

Given the experience of the pilot project (100 recordings and time-coded transcriptions totaling 38 hours; a 1,500-stem lexicon), much of the groundwork for a major documentation project has been achieved. These include, but are not limited to the following:

1. A practical orthography for the transcription of the YM recordings has been established.
2. A Fix Mixtec program has been developed by Poser than can convert the practical orthography to a IPA-based representation.
3. Amith and Castillo have successfully worked together over distance using voice-over-Internet and application sharing technology (Elluminate, which will be licensed for this project).
4. Amith and Poser have worked together on developing programs (Fix Mixtec, ShoePolish, Prompter/ Segmenter) that facilitate documentation.
5. Amith has built up a network of biologists who have committed to helping the project with their expertise.
6. A basic dictionary has been created and a detailed style sheet for YM lexicography has been developed.
7. Castillo is fully capable of using the necessary software (Shoebox and Toolbox, Transcriber, FTP transfers) and hardware (Marantz PMD670 recorder).
8. A structure for registering metadata has been created.
9. Community collaboration has been established and confirmed.

A work plan has been established to meet the following project goals:

1. Approximately 110 hours of time-coded transcriptions of endangered genres of discourse and threatened domains of cultural knowledge (to be added to the 38 hours produced in a pilot project)
2. A semantically detailed lexicon containing, at a minimum, each lemma in the corpus.
3. A concise reference grammar of YM.
4. A series of focused studies on topics of YM phonetics and phonology, and morphosyntax that are of central concern in studies of Mixtecan languages.

From its inception to conclusion the project team will carry out recording and transcribing of endangered genres of discourse and threatened domains of cultural knowledge, ethnobiological fieldwork (specimen collecting, recording of native knowledge of local flora and fauna, scientific identification of specimens), dictionary development, elaboration of a basic reference grammar. Work on all these facets of the documentation project will be steady and constant throughout the three years of the project. The amount of time spent on recording, transcription, lexicon, and grammar is given in the section on Documentation Methods.

Theoretical research: Fieldwork oriented to the theoretical issues discussed in Research Outcomes will be concentrated in years 2 and 3 of the project although discussions of methodology (e.g., on studying topic and focus constructions, on measuring nasalization, on the necessity of frames for determining lexical tone) will begin during year 1.

Parsing and preparation of transcriptions in four-line format: Poser will complete the XFST transducer in year 2, during which time it will be tested. During year 3 Poser will work with Amith and Castillo to develop the best way to run the transducer over the texts to create the gloss and parse lines of the four-line interlinear transcription, linking them to the time-coded orthographic transcription. During year 3 Castillo will work on the free Spanish translation of the transcribed materials.

Preparation of local materials: Community authorities will be consulted throughout the project, but it will be in Year 3 that practical language materials will be developed for local use.

Q16 Financial details of support requested

All costs must be given in GBP; only eligible costs will be accepted. Use http://www.xe.com for exchange rates and indicate the date of conversion.

GBP	133,170
Date	16 June 2009

All calculations are based on the following:

1 pound = 1.6352 US / 1 US = .6115 pound
1 pound = 23.0329 Mexico pesos / 1 peso = .0455 pound

Summary of budget

The figures you present here must agree with the more detailed breakdown of costs that you will provide below.

	Budget Items	**Year One**	**Year Two**	**Year Three**	**Total (GBP)**
Q16a	Replacement teaching costs	14,676	15,116	15,571	45,363
Q16b	Research staff costs	13,525	17,231	17,747	48,503
Q16c	Technical Staff	0	0	0	0
Q16d	Language Consultants	1,608	1,656	1,706	4,970
Q16e	Administrative Staff	1,529	1,529	1,529	4,587
Q16f	Student Stipend	0	0	0	0
Q16g	Equipment	180	0	0	180
Q16h	Travel and subsistence	7,184	7,184	7,184	21,552
Q16i	Consumables	2,005	2,005	2,005	6,015
Q16j	Other costs (see guidance notes for eligible items)	0	0	2,000	2,000
TOTAL		40,440	44,717	48,013	133,170

Justification of costs

Q16a Replacement teaching costs

Replacement teaching costs are for yourself and co-applicants who are in established posts and where such support is necessary to enable you to undertake extended periods of fieldwork.

Replacement teaching costs	**Total (GBP)**
Jonathan D. Amith at 33% fulltime employment over three years	£45,362
Total research staff costs (replacement teaching costs; see below for research salaries)	£45,362

Justification for replacement teaching costs/research salaries

Provide details of the basis for calculating the above replacement teaching/salary costs, their justification and the length and location of your related field trips. Official rates of pay should be used, as a comparison with the level of cost requested. You are permitted to include a reasonable estimate for annual inflation increases, please list.

Amith is an independent scholar without a teaching position or income beyond that of grant-supported activity. The level of his compensation was established as 33% of full-time employment based on his commitment to the project of 13 hours/week in addition to approximately 4 weeks/year in the field.

The figure used to calculate Amith´s prorated compensation was $72,000/year, a figure well below the salary of even entry-level assistant professors at Gettysburg College. For example, salary and compensation for a nine-month position at Gettysburg College is $82,000 (approximately $63,000 in salary and $19,000 in benefits; note that converting $82,000 for nine months to a twelve-month position yields $109,333).

Q16b Research staff costs

Specify here the total number of hours you plan to work with your research staff. You should cost these hours of work at the appropriate hourly or daily rate for the location where the staff will work. Include names, where known. You may be asked to provide evidence of the relevant pay scale(s) or hourly rate(s) used.

Name of research staff	Salary/Hourly Rate and time on the project	Total (GBP)
Research staff (1): Rey Castillo Garcia	£700/month for full-time employment, which will be for the full 36 months of the project (3% raise/year)	£25,964
Research staff (2): William Poser	£22/hour for 140 hours each in years 2 and 3 of the project (3% raise the final year)	£6,699
Research staff (3): Heriberto Avelino	No salary as Avelino is a fulltime researcher at the Max Planck Institute with a 12-month contract. He will conduct this research as part of his fulltime responsibilities	£0
Research staff (4): Monica Macaulay	Macaulay is hired on a nine-month contract at the University of Wisconsin, compensation is calculated at £185/day for 5 days/year (3% raise/year)	£2,858
Research staff (5): Pilar Mendoza	£350/month for half-time employment, which will be for the full 36 months of the project (3% raise/year)	£12,982
Total research staff costs		£48,503

Justification for research staff costs

Summarise here the duties of each of the research staff listed, their roles, responsibilities, and why they are necessary for your project.

Rey Castillo is a fluent Mixtec speaker from Yoloxóchitl, Guerrero. Castillo is key to the project as he has a masters degree in linguistics from Centro de Investigaciones y Estudios Superiores en Antropología Social, (CIESAS), Mexico, and is fully capable of accurately transcribing recordings in his native language (his masters thesis was on the phonology of YM). Over the past two years he has worked with Amith on a YM documentation project, one year of which was supported by a SOAS Pilot Project award. Castillo is not only proficient in writing YM and in analyzing its phonology and morphosyntax, but he has developed a high degree of competence in documentation best practice, from recording and the registration of metadata, to the use of software tools such as Transcriber, ELAN, Toolbox, as well as sound editing software and FTP transfers. His compensation (£700/month) is based on the present rate at the national university (UNAM) for a student with a master's degree working on an externally funded university project. Castillo will work with Amith and Avelino in the field (recordings and research), transcribe the recordings, work with Amith on the dictionary and grammar, work with Avelino and Poser on phonetic and phonological research (including obtaining primary research data). Castillo plans on pursuing a doctorate in linguistics after the completion of the project and he will use the documentation material in his future research.

William Poser is a renowned phonetician (particularly of prosodic and tonal features) and computational linguist with extensive experience in language documentation (mostly Athabaskan languages), working with indigenous communities, and developing software for documentation efforts. He has already worked with Amith in developing software (described in another section) including one, Fix Mixtec, that converts a practical YM orthography into an IPA representation. He will work closely with Avelino on the phonetic and phonological aspects of this project and with Amith on developing software to ensure that the archived material follows best practice recommendations, including the generation of 3-line interlinear formats (surface, parse, gloss) for the time-coded transcriptions. His program "ShoePolish" will be used to ensure the data integrity of the Toolbox-based lexicon and his program "Prompter/Segmenter" will be used to facilitate the elicitation of acoustic material from a database and the segmentation of the sound files that result into a formats easily manageable for research. Finally, he will be "on-call" to write programs that can facilitate the documentation process. Poser is an independent scholar with no full-time academic position. His compensation is calculated for 140 hours/year at a rate (like Amith's) of £22/hour.

Heriberto Avelino is a fulltime researcher at the Max Planck Institute. He is an expert on the phonetics and phonology of Yalálag Zapotec (an Otomanguean language), having written his doctoral dissertation ("Topics in Yalálag Zapotec, with particular reference to its phonetic structures"; UCLA 2004) on this topic. He assisted Larry Hyman in the phonetic analysis of John Daly's Peñoles Mixtec data (see Daly and Hyman, 2007). He will be responsible for developing a fieldwork agenda with Amith and Castillo to document lexical tone, to investigate questions of floating tones and sandhi, and to explore the phonetic impact of processes such as focus and topicalization. As questions emerge regarding either phonetics or phonology, he will work with Castillo, Poser, and Amith to develop a field methodology and he will be responsible for carrying out the necessary field research. He has a 12-month/year 5-year appointment at the Max Planck Institute and no salary or compensation has been included for his participation. The budget includes one round-trip airfare and in-country travel and per diem expenses for approximately 2 weeks of fieldwork recordings per year.

Monica Macaulay is one of the world's leading authorities on Mixtecan languages as well as a researcher dedicated to documentation of endangered languages (after having worked extensively on Chalcatongo Mixtec, she has been working on documenting Menominee). Her book on Chalcatongo Mixtec is arguably the best grammar on any Mixtecan language. She will be a consultant on this project to help resolve questions of best practice for the orthographic transcriptions, to review the grammar as it is being elaborated, and to develop questions or morphosyntactic analysis that might be resolved through targeted field research. She has a nine-month/year appointment at University of Wisconsin and has been budgeted for 5 days consultation/year (representing summer research compensation) at £185, about half of the accepted NSF rate.

Pilar Mendoza is a botanist who is working with Amith on comparative Nahuatl ethnobiology out of the Instituto de Biología of the Universidad Nacional Autónoma de México. Mendoza will coordinate all activities related to the collection, preservation, and identification of the biological specimens. Note, however, that field collections and determination of botanical material will be carried out pro bono by Nelly Diego and her team at the Universidad Nacional Autónoma de México (appendix, p. 7). Nevertheless, the maintenance of a database of the biological material that represents the point of departure for exegetical YM texts is extremely important to ensure the integrity of the ethnographic and ethnobiological components of the documentation project. Her compensation is £350/month and, like the full-time compensation of Castillo, reflects the official rate at the national university for researchers with a master's degree contracted for an externally funded project.

Q16c Technical staff

Specify here the total number of hours you plan to work with your technical staff. You should cost these hours of work at the appropriate hourly or daily rate for the location where the staff will work. Include names, where known. You may be asked to provide evidence of the relevant pay scale(s) or hourly rate(s) used.

Technical staff	**Total (GBP)**
Note that Amith and Poser (member of the Linguistic Society of America Technology Advisory Committee 2008–present) are proficient in all the technical skills necessary for best practice documentation and archiving procedures.	0
Total technical staff costs	0

Justification for technical staff costs

Summarise here the duties of the technical staff, their roles, responsibilities, and why they are necessary for your project.

Amith and Poser each has over a decade of experience in language documentation projects. Amith will, as part of his responsibilities, ensure that best practices are followed in the field in terms of ethical and professional recordings. He will collaborate with Castillo on the technical issues of orthographic conventions for transcriptions and the elaboration of a detailed metadata XML file.

Poser is an expert programmer and will be responsible, in particular, for computational and technical support to the project.

In both cases, however, Amith and Poser are considered part of the research, not technical, staff.

Q16d Language consultants

Specify here the total number of hours you plan to work with language speaker consultants. You should cost these hours of work at the appropriate hourly or daily rate for the location where you will be doing your research. Indicate how you have calculated this pay rate. Include names of the consultants if known.

Language consultants	Total (GBP)
Rey Castillo, a native speaker, is considered a member of the research staff and not a language consultant. He will, nevertheless, be the key language consultant for the development of the YM dictionary/lexicon, with Amith in an advisory and collaborative position on lexicography and electronic text development. The language consultants per se (i.e., those contracted to provide digitally recorded textual materials) will be divided into two major groups. The first will be those who provide narrations, stories, and exegetical texts on quotidian activities for corpus development. Calculations are for 50 days of consultation at £12/day per person. Generally, two consultants will be hired to record together for a day, taking alternating turns before the microphone, to avoid exhaustion and ensure better recordings. The calculation, therefore, is for 25 days of fieldwork in recording narrations, stories, and exegetical texts. Total cost is £600 for the first year, with 3% increase over the following two years. The second major group will be those experts on the natural environment who will help in ethnobiological research by documenting nomenclature and providing exegetical texts of natural phenomena. In this case, given the wide variation in expertise among speakers, special care will be taken to choose three who have been found to be both knowledgeable and highly skilled speakers. These three individuals will work in a team with Amith, Castillo, and a pair of biologists. Calculations are for 28 days of fieldwork, with the language consultants receiving at £12/day per person. Total is at £1008 for the first year, with 3% increase over the following two years. The rate of £12/day is based on the salary of a skilled worker in the research area.	£1855 £3115
Total language consultant costs	£4970

Justification for language consultant costs

Summarise here the duties of the language consultants, and their roles in the project.

The two groups of language consultants are essential to the success of the documentation effort.

The first group will be varied in age, sex, and social status and the 50 days of consultation might be distributed among 25–35 different individuals. The goal of working with this group it to obtain a highly varied set of material from persons with quite varied knowledge of different endangered genres of discourse and threatened domains of cultural knowledge.

A day of recording with two individuals might produce 2–3 hours of digital material. Thus 25 days of recording will yield between 50 and 75 hours of material. The content will vary but the relative abundance of recordings will mean that the effort of transcription will be applied only to the most deserving digital recordings. All recordings will be archived, however, even though the project will select 110 hours of the recordings for time-coded transcription.

The second group, that working on the exegetical texts on the natural environment, will be more constant as the individuals will be selected for their knowledge of this particular domain. The team of three individuals, once selected, will accompany a group of biologists throughout the day. It is estimated that the 28 days of planned fieldwork per year will produce approximately 42 hours of digital recording. The thematic focus of the texts will be not only on the natural environment, but on the production of material objects used in the culture.

Thus the present project will produce approximately 250–300 hours of digital recordings of which 110 hours will be selected for time-coded transcription. All recorded material will be archived with fully elaborated metadata in an XML database.

Q16e Administrative staff

Specify here the total number of hours you plan to work with administrative staff. You should cost these hours of work at the appropriate hourly or daily rate for the location where you will be doing your research. Indicate how you have calculated this pay rate. You may be asked to provide evidence of the relevant pay scale(s) or hourly rate(s) used. Include names of the staff if known.

Administrative staff	**Total (GBP)**
Sharon Kuhn, Senior Accountant, Financial Services Office, Gettysburg College	£4,586
John Ryan, Vice-Provost	0
Total administrative staff costs	£4,586

Justification for administrative staff costs

Summarise here the duties of the administrative staff, and their roles in the project.

The administrative staff of Gettysburg College will be headed by John Ryan (Vice-Provost) and Sharon Kuhn. The project will be administered directly out of the vice-provost's office, which will be responsible for ensuring that all research requirements are met, particularly in regard to office space, supplies, interlibrary loan, and secretarial support. The vice-provost's office will carry out this effort probono.

Sharon Kuhn, senior accountant, will be responsible for accounting and administration of the financial aspects of the SOAS grant, should it be awarded. Her total time on the project is estimated to be 36 hours/year or 108 hours over the three-year duration. Her time, which includes support by other members of the Financial Services Office, has been calculated at £42.5 pounds/hour. She will ensure that all honoraria and salaries are promptly paid, that other expenses are reimbursed or, if major, funded through advances provided by the college.

It should be noted that the cost for grant administration is only £4,586. This represents approximately 3.5% of the total budget. This extremely low figure reflects Gettysburg College's commitment to supporting privately funded grant initiatives in areas, such as language documentation, in which the funds are urgently needed for the project itself.

Justification of costs

Q16f Stipend

The stipend is to cover the living cost of students at the host institution. Ineligible categories of costs are: tuition fees, other administrative costs at the host institution, extra income to cover income tax.

Stipend	**Total (GBP)**
Year 1	0
Year 2	0
Year 3	0
Total stipend costs	0

Justification: Provide breakdown of your stipend costs

Q16g Equipment costs

You may apply for equipment to be used for your project. Provide specific details of the equipment you plan to purchase and its cost. Indicate how you calculated the cost(s).

Specification	Quantity	Unit Cost (GBP)	Total (GBP)
External hard drives (2 x 500 gigabytes) for backup	2	£90 each	£180
Total equipment costs			£180

Justification for equipment

Provide full justification for all your equipment and its role in meeting your project's goals. Also list here other equipment that you will be using on the project, which is not included in your budget proposal.

No further equipment purchases are foreseen at this time beyond two 500 gigabyte external hard drives for data storage and backup. Through the previous ELP Pilot Project awarded to Amith as well as his other documentation initiatives, Amith has built up a significant collection of equipment necessary for proper documentation:

- Marantz PMD 670 digital recorder (1 unit)
- Sonifex Courier digital recorder (1 unit)
- Audio Technica ATM 75 headworn unidirectional mikes and XLR cables (4 units)
- Sony HD HVR A1U video (high definition) camera and shotgun Sennheiser microphone
- Canon EOS 20D digital camera with macro, telephoto, and wide-angle lenses
- Bogen-Manfrotto and Libec (Th 650dv) tripods and heads for digital still and video photography
- Amith maintains unlimited server space for online back-up of all materials before formally archived at ELP (as the first 100 recordings/38 hours has been)
- Batteries for all recording units (digital audio, high definition video, still)
- Flash cards for digital recording
- Materials for ethnobiological research (plant presses and dryers, GPS units, storage units

Note that Heriberto Avelino and, through him, the rest of the project team will have access to the equipment (stationary and field) of the phonetics lab at the Max Planck Institute. This includes a Nagra VI six-channel digital audio recorder as well as specialized equipment for phonetic recordings.

Q16h Travel and subsistence costs

You may apply for the cost of travel and subsistence for fieldwork. Provide specific details of your travel and subsistence costs.

Travel and subsistence	Total (GBP)
International airfare for Amith (US-Mexico): 4 trips/year at £480 each trip	5,760
International airfare for Heriberto Avelino (Netherlands-Mexico): 1 trip/year at £812	2,436
Country internal travel: £1260/year	3,780
Hotels in Mexico for Amith and Avelino before getting to the field site (where stays will be with families): 11 days/year for both at £36/night x 3 years	1,188
Per diem for Amith and Avelino in Mexico: 60 days total each year at £20/day	3,600
Per diem for ethnobiological fieldwork in the field: team of 5 individuals in addition to Amith: £57/day for 28 days/year	4,788
Total travel and subsistence costs	21,552

Justification for travel and subsistence costs

Provide full justification for all your travel and subsistence costs.

Airfare is calculated based on present costs: Amith is budgeted for four trips to the research area (for a total of 4-6 weeks of fieldwork/year) and Avelino is budgeted for one trip of approximately 2 weeks. Amith will split his time between working with Castillo and the two groups of native language consultants. Thus he will work for 2–3 weeks with the first group responsible for providing narrations, stories, and exegetical texts on quotidian activities for corpus development. He will also work for 2–3 weeks with the second group, experts on the natural environment who will help in ethnobiological research by documenting nomenclature and providing exegetical texts of natural phenomena. During this time he will be accompanied by a team of botanists from Nelly Diego's lab at the Universidad Nacional Autónoma de México.

It is estimated that each year Amith and Avelino will need 11 nights at hotels in Mexico City and Acapulco in their trips to the field location (about 3 hours from Acapulco). The cost of £36/night reflects a present costs in Mexico City and Acapulco.

In country travel is estimated as compensation for gas and tolls for Amith (who has a vehicle in Mexico) and bus fare for Avelino.

Finally, per diem is £20/day in urban areas and about half that for stays in the field.

Q16i Consumables

You may apply for the cost of consumables (e.g. recording media, postage, communications, paper, fuel for vehicles or generators) directly associated with the project. Note that ELDP does not support the following costs: overheads, direct central administration, indirect costs or renting office accommodation in the field.

Specification	Quantity	Unit Cost (GBP)	Total (GBP)
Preparation (mounting) of botanical and arthropod specimens; acquisition of vials, alcohol, newspaper, cardboard	Yearly estimate	875/year	2,625
Mailings (particularly of biological specimens for determination)	Yearly estimate	270/year	810
Internet connection for Rey Castillo	36 months	19/month	675
Licensing of software (Elluminate) for application sharing and conferencing	Yearly license	635/year	1,905
Total consumable costs			6,015

Justification of consumable costs

Provide full justification for all your consumable costs.

A significant aspect of this project is its effort to document, through recordings, endangered domains of cultural knowledge. One important area of concern is ethnobiological expertise. To ensure that the exegetical texts on natural history are accurately linked to specimens that are determined (i.e., identified by scientific nomenclature) by experts, the material needs to be prepared and shipped.

The first two items represent estimated costs of preparing and preserving approximately 250 plant and 250 arthropod specimens/year. It is based on past experience. The materials need to be sent to experts for determination and, again based on past experience, approximately £270/year will be needed.

The remaining two budget items represent the cost of maintaining daily contact between Amith and Castillo. First, Castillo needs to be provided with a high-speed Internet connection, here calculated at £19/month for 36 months.

Likewise, Amith and Castillo (as well as other members of the project) need to work collaboratively on electronic texts online. Amith has been licensing a program, Elluminate, for over three years. This program permits individuals to share applications. For example, in the Pilot Project, Castillo would send the lexicon, or a transcription, to Amith, who would open it up on his computer. Castillo is able to see this document and edit it. Or, Castillo would show Amith a particular text and Amith could comment on it or edit it, while both users were observing the same screen. The cost of the license is £635/year.

Q16j Other costs

Other costs, as permitted by the Guidelines for Applicants.

Itemised expenditure	Total (GBP)
Preparation of practical language materials for the indigenous community	2000
Total consumable costs	20

Justification of other costs

Provide justification of how these items will support the project's goals.

Toward the end of the documentation project, Amith and Castillo will consult with the Yoloxóchitl community to determine the best way in which to transform the project's results into practical materials.

For example, in his support letter for Amith and Castillo, Fernando Nava, the director of the Institución Nacional de Lenguas Indígenas, refers to the institution's interest (pending funding availability) of producing for native-speaking communities the same sort of work in Yoloxóchitl Mixtec that the INALI edited (with a press run of 10,000) for Amith in Nahuatl: a set of 6 CDs with 6 hours of digital recordings in the native language accompanied by a book of transcriptions prefixed by a brief grammatical sketch. Should this become possible, the £2,000 could be used to help prepare (e.g., design) the final manuscript.

The support letter from the Yoloxóchitl authorities specifically mentions interest in ethnobotanical and ethnozoological topics. Amith is working with an illustrator on practical language materials in Nahuatl on local insects and spiders (illustrations accompanied by native language texts). The same illustrator could be contracted to prepare a book on YM knowledge of local flora and fauna and the £2,000 used to defray the costs of illustration and design.

Q17 Graduate studentship details

Provide details of students whose stipend costs you are applying for.

	Student 1	Student 2
Full name of graduate student to be supported by the grant		
Details of any funding for postgraduate study		
Institution and department where registered		
Nature and scope of dissertation		
Date registered for postgraduate study		
Proposed submission date of dissertation		
Supervisor’s title, full name and department		
Supervisor’s comments: please comment briefly on the student’s qualifications and experience		

Q18 Referees

Supply full contact details of two referees who will write in support of your application. You must send a completed copy of this application plus a referee form to each of your referees. Your referees must return the completed referee form by email attachment direct to the ELDP office, at eldp@soas.ac.uk. The Programme will also ask for comments from independent referees.

Referee details (1)

Name	William Merrill
Position	Curator (equivalent of full professor) Department of Anthropology National Museum of Natural History, Smithsonian Institution
Address	10th Street and Constitution Avenue, NW Washington, D.C. 20560
Telephone Number	In Mexico City 52-55-5207-0406
Email	ralamuli@gmail.com or merrillw@si.edu

Why have you chosen this referee?

William Merrill is the founder of the Mexico-North Research Network, where I am the director of the initiative on Endangered Languages. He and I have known each other for years and he is well aware of my work on documenting endangered languages and cultures.

Merrill is fluent in Tarahumara (a Southern Uto-Aztecan language spoken in the state of Chihhuahua) and has worked closely with indigenous individuals and groups in a project on biodiversity and indigenous knowledge. He himself has been a force in getting the Smithsonian Institution to recognize the importance of documenting endangered languages and cultures and played a key role in establishing documenting endangered languages as a recently approved priority for Smithsonian Institution initiatives. In his role of curator of the North American collection, Merrill has been at the forefront of repatriation efforts. He has worked closely with indigenous groups of the United States southwest in negotiating the return of cultural objects and is keenly sensitive to the ethical issues involved in language documentation.

Finally, through his work on biodiversity in the Tarahumara, Merrill has come face-to-face with the rapidly declining knowledge of the natural environment that has affected indigenous peoples of the Americas and has stressed (in the biodiversity project and at the Smithsonian) the importance of combining ethnography, particularly ethnobiology, with language documentation.

Referee details (2)

Name	Michael Maxwell
Position	Area Director for Technology Center for Advanced Study of Language University of Maryland
Address	7005 52nd Avenue College Park, MD 20742
Telephone Number	301-226-8895
Email	maxwell@umiacs.umd.edu

Why have you chosen this referee?

Michael Maxwell is one of the world's leading authorities on the application of natural language processing to language documentation efforts and has worked with Amith on the development of a Nahuatl language transducer (capable of both parsing texts and generating surface forms from glosses or underlying representations) that will greatly facilitate generating four-line interlinear format from transcriptions of digital recordings.

Maxwell and Amith met in 2001 at the Linguistic Data Consortium, University of Pennsylvania, where Maxwell was a senior programmer and Amith a visiting researcher. They began to discuss the implications of natural language processing for archiving endangered language materials. Besides their collaboration on the Nahuatl transducer, they have coauthored two articles on strategies for archiving executable morphological grammars and its relevance for permanent archiving of endangered languages (see Amith CV: in press; 2005a).

Maxwell has served on the NSF panel for their Documenting Endangered Languages initiative and has been an active participant in the E-MELD project. He is extremely well aware of the importance of following best practices in endangered language documentation and of the need to endure permanency in the archival record.

He will be able to comment on Amith's work and on the importance of following best practice in endangered language documentation and archiving.

Q19 Curriculum Vitae

Using this template, provide a CV for the applicant and every person in the budget, with the exception of language consultants. Include all CVs as part of the overall application form; do not submit them separately, you can append additional CV forms to the application and these can be found at www.hrelp.org/grants/apply.

Major Documentation Project

Curriculum Vitae

First Name	Jonathan	Title, if any	Dr.
Family Name	Amith	Nationality	USA
Title of current post or study	Independent researcher	Date of appointment or registration	n.a. (working independently on different grants/projects since receiving Ph.D.)
Name or employer or place of study	Research fellow Gettysburg College, Research affiliate Smithsonian Institution, Research associate University of Chicago. Present grant support: National Science Foundation, "Nahuatl Language Documentation Project (NLDP): Sierra Norte de Puebla"; 39% full-time employment		

Education/Training

List your highest/latest qualification first

Dates of study	Degree	Subject	University/Institution
1980–2000	Ph.D.	Anthropology	Yale University
1970–77	Bachelor's degree	General Studies	University of Michigan, Ann Arbor

Employment

List your last 3 positions (Please note that since graduating with a Ph.D. in anthropology from Yale University in 2000, I have been supporting myself through grants as an independent scholar working on documentation of endangered languages)

Date	Position	Name of Employer
Aug. 2008–	Principal investigator	NSF Grant for Documentating Endangered Languages: "Nahuatl Language Documentation Project: Sierra Norte de Puebla."
Aug. 2005–Aug. 2008	Principal investigator	NSF Grant for Documentating Endangered Languages: "Guerrero Nahuatl Language Documentation and Lexicon Enrichment"
Jan. 2001–Aug. 2005	Principal investigator	Various grants from the US Department of Education, the National Endowment for the Humanities, and the Ford Foundation, all on the Nahuatl language and culture

List any other significant awards, exhibitions, or other achievements relevant to this application

Books

2009 (ed.) *Ok nemi totlahtōl, vol. 1, Estado de Guerrero: San Agustín Oapan, Ameyaltepec, San Francisco Ozomatlán.* Mexico City: Instituto Nacional de Lenguas Indígenas.

2005 *The Möbius Strip: A Spatial History of Colonial Society in Guerrero, Mexico*. Stanford University Press.

1995 (ed.) *The Amate Tradition: Innovation and Dissent in Mexican Art.* Chicago: Mexican Fine Arts Center Museum; Mexico City: La Casa de las Imágenes [distributed by University of New Mexico Press] Awarded special recognition for the category non–black-and-white book by the Unión de Industriales Litógrafos de México, A.C.

Selected articles and book chapters

in press "Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language (senior co-author Susan G. Guion; junior co-authors Christopher S. Doty and Irina A. Shport). *Journal of Phonetics*.

in press "Language Documentation: Archiving Grammars." *Chicago Linguistic Society*, vol. 41 (senior co-author Michael Maxwell).

2009 "La fonología y la escritura en los pueblos de habla náhuatl en la cuenca del Río Balsas Alto, Guerrero: Ameyaltepec, San Agustín Oapan y San Francisco Ozomatlán." In Jonathan D. Amith, ed., *Ok nemi totlahtōl, vol. 1, Estado de Guerrero: San Agustín Oapan, Ameyaltepec, San Francisco Ozomatlán.* Mexico City: Instituto Nacional de Lenguas Indígenas.

2005a "Language documentation: The Nahuatl grammar." In *Computational Linguistics and Intelligent Text Processing: 6th International Conference, CICLing2005, Mexico City, Mexico, February 13–19, 2005. Proceedings*, ed. Alexander Gelbukh. Berlin: Springer, pp. 474–85 (senior co-author Michael Maxwell).

2005b "Place making and place breaking: Migration and the development cycle of community in colonial Mexico." *American Ethnologist* 32(1): 159–79.

2002 "What's in a word? The *why's* and *what for's* of a Nahuatl dictionary. In *Dictionaries of Indigenous Languages of the Americas,* eds. William Frawley, Kenneth Hill, and Pamela Munro. Berkeley:

University of California Press, pp. 219–58.

1998 “Tan ancha como tu abuela: adivinanzas en náhuatl del Guerrero central.” *Tlalocan,* 12: 141–219.

1995a “The creation of indigenous images: from private nightmares to political protest. In Jonathan D. Amith (ed.), *The Amate Tradition.* Chicago; Mexico City: Mexican Fine Arts Center Museum; La Casa de las Imágenes, pp. 41–100.

1995b “The history of the Balsas River basin Nahua communities.” In Jonathan D. Amith (ed.), *The Amate Tradition.* Chicago; Mexico City: Mexican Fine Arts Center Museum; La Casa de las Imágenes, pp. 129–44.

1994a “Transitive nouns and split possessive paradigms in Central Guerrero Nahuatl. *International Journal of American Linguistics* 60: 342–68 (co-author Thomas Smith-Stark).

1994b “Predicate nominal and transitive verbal expressions of interpersonal relations *Linguistics* 32: 511–47 (co-author Thomas Smith-Stark).

1988 “The use of directionals with verbs in the Nahuatl of Ameyaltepec, Guerrero.” In J. Kathryn Josserand and Karen Dakin (eds.). *Smoke and Mist: Mesoamerican Studies in Memory of Thelma D. Sullivan.* Oxford: B.A.R., vol. 2, pp. 395–421.

Funded research on language documentation and community collaboration

2008–11	Nahuatl Language Documentation Project: Sierra Norte de Puebla. National Science Foundation, Documenting Endangered Languages ($291,798, Award #0756536)
2008–9	Corpus and lexicon development: Endangered genres of discourse in Tu̲’un ísa̲ví (Mixtec) of Yoloxóchitl, Guerrero. Hans Rausing Endangered Languages Project, School of Oriental and African Studies (£ 5,345 / $10,364)
2007–10	Ford Foundation grant ($89,000 Donativo 1075-0256): “Las Lenguas Indígenas de la Montaña (Estado de Guerrero)”
2006–9	Ford Foundation grant ($49,464; Donativo 1065-0692): “Community and School Outreach for Nahuatl Education and Literacy”
2005–8	National Science Foundation/National Endowment for the Humanities. Documenting Endangered Languages ($299,917; Award #0504164): “Guerrero Nahuatl Language Documentation and Lexicon Enrichment”
2004	Foundation for the Advancement of Mesoamerican Studies ($9,800; FAMSI #03049): “Nahuatl Cultural Encyclopedia: Botany and Zoology”
2003–6	Yale University Title VI program included approximately $75,000 over 3 years for the development of educational tools for Nahuatl (participating research)
2002–6	Ford Foundation grant ($70,000; Donativo 1025-1428): “Nahuatl Language Documentation and Literacy”
2002	National Endowment for the Humanities Fellowship ($40,000 stipend; FB-37654) for “A Cultural Encyclopedia for the Nahuatl Learning Environment”
2002	Title VI International Research Scholarship Program, Department of Education grant ($58,500; Award #P017A010051) to support the electronic publication of “The Nahuatl Learning Environment”
2001–2	Foundation for the Advancement of Mesoamerican Studies ($4,700): “A Cultural and Pedagogical Lexicography of Modern Nahuatl”
2001–2	Fund for Culture Mexico/USA (auspiced by the Rockefeller Foundation, the Fundación Cultural Bancomer and the Fondo Nacional para la Cultura y las Artes; $20,000 grant to fund the Nahuatl Transcultural Learning Project*)*
2000–2	Title VI International Research Scholarship Program, Department of Education ($98,958; Award #P017A990031) to support the elaboration of “The Nahuatl Learning Environment” / lexicon
1994	Fund for Culture Mexico/USA (auspiced by the Rockefeller Foundation, the Fundación Cultural Bancomer and the Fondo Nacional para la Cultura y las Artes; $25,000 subvention for the book *The Amate Tradition: Innovation and Dissent in Mexican Art)*

Books relevant to language documentation:

Editor of *Ok nemi totlahtōl, vol. 1. Estado de Guerrero: San Agustín Oapan, Ameyaltepec, San*

Francisco Ozomatlán. Mexico City: Instituto Nacional de Lenguas Indígenas. (collection of 6 CDs or oral traditions and monolingual transcription in Nahuatl, for use in native speaking communities and schools; published 2009)

Websites and online material:

Nahuatl Learning Environment: Online Nahuatl dictionary
http://www.balsas-nahuatl.org & http://nahuatl.ldc.upenn.edu (Username: oapan Password: nahuatl)

Nahuatl Language Documentation
http://www.balsas-nahuatl.org/documentation

Yoloxóchitl Mixtec Pilot Project
http://www.balsas-nahuatl.org/soas (sound files and transcriptions from the pilot project on YM; archived at ELDP)

Software development by various programmers in association with Amith
http://www.balsas-nahuatl.org/toolset

Collaborations:

Individual

Mike Maxwell: computational morphology and the development of a transducer (parser) for Oapan Nahuatl (two co-authored articles)

Susan Guion: tonogenesis in Balsas Nahuatl (one co-authored article)

Botanists and zoologists listed in appendix (approximately 100): for scientific determinations relevant to ethnobiological research (article on Nahuatl ethnoentomology in preparation and book being developed)

Bill Poser: development of software for language documentation; Poser programmed the following: ShoePolish, Prompter/Segmenter, Fix Mixtec. For a description of these programs, see Poser CV

Institutional

Tosepan Titataniske: Work with indigenous collective to set up an indigenous controlled museum and herbarium to house the results of all documentation and other research in which Tosepan is involved, making the results available to collective members, the Tosepan Montessori schools for the children of its members, indigenous people of the region and beyond, and outside scholars and students.

Exhibitions

Curator for *The Amate Tradition: Innovation and Dissent in Mexican Art* [Exhibition of paintings, ceramics, carved wood, ethnographic objects and photography from the region of San Agustín Oapan, Guerrero] Mexican Fine Arts Center Museum, Chicago; Jan. 27–May 27, 1995. Editor and major contributor to catalogue/book *The Amate Tradition: Innovation and Dissent in Mexican Art.* Chicago: Mexican Fine Arts Center Museum; Mexico City: La Casa de las Imágenes.

Major Documentation Project

Curriculum Vitae

First Name	Rey	Title, if any	
Family Name	Castillo	Nationality	Mexican
Title of current post or study	Masters degree in lingüística indoamericana	Date of appointment or registration	
Name or employer or place of study	Centro de Investigaciones y Estudios Superiores en Antropología Social		

Education/Training

List your highest/latest qualification first

Dates of study	Degree	Subject	University/Institution
2005-2007	Master´s degree in lingüística indoamericana	La fonología segmental y tonal del mixteco de Yoloxóchitl, Guerrero	Centro de Investigaciones y Estudios Superiores en Antropología Social
1996-2000	Bachelor's degree in education	La formación práctica y docente del el enfoque intercultural	Universidad Pedagógica Nacional

Employment

List your last 3 positions

Date	Position	Name of Employer
2008–9	Department of Training in (native speakers)	Instituto Nacional de Lenguas Indígenas
2002–5	Teacher of migrant children	Secretaría de Educación Pública
1993–2002	Teacher	Secretaría de Educación Pública

List any other significant awards, exhibitions, or other achievements relevant to this application

Master's Thesis

2007. "Descripción fonológica segmental y tonal del mixteco de Yoloxóchitl, Guerrero." Thesis para Maestría en Lingüística Indoamericana, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), Mexico City, Mexico.

Thesis Director: Dr. Thomas C. Smith Stark
Readers: Dr. Esther Herrera Zendejas, Mtro. Francisco Arellanes

Bachelor's Thesis

2001 "La formación y práctica del docente indígena desde el enfoque intercultural". Universidad Pedagógica Nacional (Unidad Ajusco), Mexico City, Mexico.
Thesis Committee: Mtra. Gisela V. Salinas Sánchez, Mtra. Elena Cárdenas Pérez, Mtro. Sergio Monge Rico, Mtra. Lara Elena Ayala Lara.

Scholarships

2005–7 CONACyT scholarship for a master's degree problem in Lingüística Indoamericana (CIESAS).
1996–2000 Scholarship from the Secretaría de Educación, State of Guerrero, for bachelor's degree studies in indigenous education.

Documentation:

2007–9 Research on the documentation project "Corpus and lexicon development: Endangered genres of discourse in Tu̲'un ísa̲ví (Mixtec) of Yoloxóchitl, Guerrero" supported by the Hans Rausing Endangered Language Project (Pilot Project) and the Ford Foundation.
Responsibilities: Field recordings (totaling 100 items/38 hours); metadata creation (XML format); time-coded transcription (Transcriber format); lexicon/dictionary development (in Toolbox)

Major Documentation Project

Curriculum Vitae

Using this template, provide a CV for the principal applicant. Depending on the grant type, you may be required to submit a similar CV for other members of the team; see the relevant guidelines for details. Include all CVs as part of the overall application form; do not submit them separately, for further CV forms go to www.hrelp.org/grants/apply.

First Name	William	Title, if any	Dr.
Family Name	Poser	Nationality	USA & Canada
Title of current post or study	self-employed freelance linguist and programmer	Date of appointment or registration	

Education/Training

List your highest/latest qualification first

Dates of study	Degree	Subject	University/Institution
1979-1985	Ph.D.	Linguistics (with Electrical Engineering)	MIT
1974-1979	A.B.	Linguistics (with Classics)	Harvard

Employment

List your last 3 positions

Date	Position	Name of Employer
2001–5	Lecturer in Linguistics	University of Pennsylvania
1994–98	Associate Professor (First Nations Studies)	University of Northern British Columbia
1983–94	Assistant Professor, then Associate,	Stanford University

	Professor with tenure (Linguistics)	

List any other significant awards, exhibitions, or other achievements relevant to this application

Committee membership relevant to documentation

Linguistic Society of America Technology Advisory Committee 2008-present
Member of the EMELD (Electronic Metastructure for Endangered Languages Data) Ask-an-Expert Panel. September 2004 to present.

Employment history

Research Associate, Linguistic Data Consortium, 2001–6

Acting Executive Director, Yinka Dené Language Institute, 1999–2001

Education Technical Advisor, Carrier Sekani Tribal Council, 1999–2000
Lheidli T'enneh Linguist Nov. 1998–Aug. 2001
Adjunct Professor, Linguistics, University of British Columbia 1999–present
Lecturer University of Pennsylvania, 2001–5
Faculty member LSA Summer Institute 1987, 1989, 1995
Visiting Erskine Fellow, University of Canterbury, Christchurch, New Zealand 1995
Visiting Researcher, ATR Automatic Interpreting Telephony Laboratory, Osaka, Japan. 1988–89

Documentation efforts

Extensive field research on and community work with the Carrier language of British Columbia since 1992. Results include dictionaries of Stuart Lake, Stony Creek, Lheidli and Cheslatta dialects, grammatical sketches of Stuart Lake and Lheidli dialects, a first year university textbook for the Stony Creek and Lheidli dialects, a short (50pp.) book aimed at lay people, and a nearly completed scholarly book on the language, as well as numerous talks and papers. I have taught Carrier (Stuart Lake and Stony Creek dialects) at university level. I have also taught introductory linguistics and lexicography courses for Carrier and Chilcotin students.

Academic research and publications

Books

Poser, William J. 2009. *An Introduction to the Carrier Language: Lheidli Dialect*. Prince George, British Columbia: Lheidli T'enneh and College of New Caledonia. [First year university-level textbook.]

Poser, William J. 2008. *An Introduction to the Carrier Language: Saik'uz Dialect*. Vanderhoof, British Columbia: Saik'uz First Nation and College of New Caledonia. [First year university-level textbook.]

Campbell, Lyle, and William J. Poser. 2008. *Language Classification: History and Method*. Cambridge: Cambridge University Press.

Poser, William J. 2008. *Saik'uz Whut'en Hubughunek* (Stoney Creek Carrier Lexicon). Vanderhoof, BC: Saik'uz First Nation. Fifth edition.

Poser, William J. 2001. *Lheidli T'enneh Hubughunek* (Fort George Carrier Lexicon). Prince George, BC: Lheidli T'enneh. (May 2001) Third edition.

Poser, William J. 1998. *Nak'albun/Dzinghubun Whut'enne Bughuni* (Stuart/Trembleur Lake Carrier Lexicon). Vanderhoof, British Columbia: Yinka Dené Language Institute. Second edition.

Poser, William J. 1984. "The Phonetics and Phonology of Tone and Intonation in Japanese." Ph.D. dissertation, Massachusetts Institute of Technology. [Includes an instrumental study that established what is essentially our current understanding of phrasal trends in F0 in Japanese.]

Selected Articles

Poser, William J. 2005. “Noun Classification in Carrier,” *Anthropological Linguistics* 47:143–68.

Trout, J. D., and William J. Poser. 1990. “Voicing, Phonotactics, and Place: Auditory and Visual Influences on Phonemic Restoration under Complementary Sentential Conditions,” *Language and Speech* 33.123–37.

Poser, William J. 1990. “Word-Internal Phrase Boundary in Japanese,” In S. Inkelas and D. Zec, eds., *The Phonology-Syntax Connection*. Center for the Study of Language and Information, Stanford University and University of Chicago Press. pp. 279–87.

Poser, William J. 1990. “Evidence for Foot Structure in Japanese,” *Language* 66:78–105. Reprinted in Natsuko Tsujimura, ed., *Japanese Linguistics: Critical Concepts in Linguistics*. Oxford: Routledge, 2005, pp. 159–90.

Poser, William J., and Yoshinori Sagisaka. 1989. “Modelling Phrase Level F0 Phenomena in Japanese,” *Densi Zyoohoo Tuusin Gakkai*, Preprint SP88-160. 24 March 1989.

Poser, William J. 1982. “Phonological Representation and Action-at-a-Distance,” In H. van der Hulst and N.R. Smith, eds., *The Structure of Phonological Representations*. Dordrecht: Foris. pp. 121–58.

Programming

Three decades experience as a programmer and Unix system administrator. Particular expertise in tools for linguistic research and text processing, regular expressions and related pattern matchers, character encoding and Unicode, sorting, digital signal processing, and phonetics software. Greatest expertise in C, Awk, and Tcl, but experience with over thirty programming languages. Experience includes the generation of Tex from Shoebox-style lexical databases for my own Carrier dictionaries, Sally Thomason's Flathead dictionary, and Jonathan Amith's Oapan/Ameyaltepec Nahuatl dictionary.

Programs developed as part of Jonathan Amith’s language documentation initiatives in Nahuatl and Mixtec.

ShoePolish – a maintenance tool for lexical databases of the Shoebox type. It can perform a number of global changes, such as assigning unique Ids to databases that lack them, renaming tags, and splitting fields containing multiple values or merging multiple single-valued fields into a single multi-valued field and can convert from one Shoebox- or Toolbox-type format to another or between CSV format and Shoebox/Toolbox-type formats. It also provides searches using arbitrary Boolean expressions over pairs of regular expressions for tags and a choice of regular expression, number predicate, or date predicate for values (e.g. “find all records in which either the Ameyaltepec or Oapan word begins with an *m*, there is an associated pitch track, the word is not a compound, and the record was last modified prior to 2006.”). Finally, it can select records that are suspicious in any of a number of respects, e.g. lacking an obligatory field or containing only one of two fields that should only occur paired. It can generate a histogram of tags so that the user can inspect records containing likely to be erroneous rare tags.

Prompter/Segmenter – a program that reads a list of words in any of several formats including Shoebox/Toolbox-type formats, filters the list according to specified criteria and sorts it in any of several ways, and generates prompts containing up to three specified pieces of information. When a recording sequence begins, a special tone is emitted for use in synchronizing time stamps. Each time the user presses the “Next” button, a new prompt is emitted and a time stamp is recorded.

Fix Mixtec – a program that takes as input a plain ASCII transcription of Mixtec and converts it to a Unicode-encoded transcription with superscript tone numbers and marking of nasalization.

Other major software programs.

Mort – a sort program that permits parsing into records other than lines, identifying key fields by tag, specifying a separate sort order for each key, and key-specific exclusions and substitutions. For each key comparisons may be lexicographic, numeric, numeric string, hybrid, by string length, by angle, by date, by domain name, by time, by ISO8601 date/time stamp, by month name, or random. It handles optional keys and provides Unicode normalization and full case-folding. Numeric, numeric string, and hybrid keys may use any of dozens of numeral systems. Reviewed at *linux.com*: http://www.linux.com/feature/134956

Libuninum – a for converting between integers and numerals in 70 numeral systems including 14 Chinese/Japanese variants, Devanagari, Kharosthi, Old Persian, and Klingon.

Redet – a tool for developing and executing regular expressions, with support for over 50 programs. Reviewed in *Linux Magazine*: http://www.linux-magazine.com/issue/65/Free_Software_Projects.pdf

SndBite – a specialized audio editor designed for breaking large recordings into small pieces (e.g. one utterance per file) with maximum efficiency. Special features include: (a) multiple simultaneous views of the waveform at different resolutions; (b) the ability to position window edges at transitions between sound and silence; (c) automated setting of cut points at zero-crossings; (d) automatic filename generation easily controlled by the user. A derivative under development called Segmenter adds the ability to read and write segment lists and to play back or write out the audio in a segment list, features for editing segment lists, and the ability to read Prompter journal files and to detect the synchronization tone generated by Prompter and use it to normalize time stamps.

Computational course relevant to documentation efforts

Developed and taught course on “Computational Methods in Linguistic Research” for University of Pennsylvania. 2002.

Editorial Positions

Founding co-editor Northwest Journal of Linguistics 2007 to present
Associate Editor, Language, 1991-1993

Many publications and much software are available at: http://billposer.org.

Major Documentation Project

Curriculum Vitae

Using this template, provide a CV for the principal applicant. Depending on the grant type, you may be required to submit a similar CV for other members of the team; see the relevant guidelines for details. Include all CVs as part of the overall application form; do not submit them separately, for further CV forms go to www.hrelp.org/grants/apply.

First Name	Heriberto	Title, if any	Dr.
Family Name	Avelino	Nationality	Mexican
Title of current post or study	Senior researcher, Max Planck Institute for Evolutionary Anthropology	Date of appointment or registration	March, 2009

Education/Training

List your highest/latest qualification first

Dates of study	Degree	Subject	University/Institution
1998–2004	Ph.D.	Linguistics	UCLA
2004–6	Post-doc	Linguistics	UC Berkeley

Employment

List your last 3 positions

Date	Position	Name of Employer
2008	Adjunct Professor Linguistics	University of Toronto
2007	Assistant Professor	Stanford University
2006	Adjunct Assistant Professor	University of California Berkeley

List any other significant awards, exhibitions, or other achievements relevant to this application

Additional Employment history

Visiting Scholar. Department of Linguistics. University of California, Berkeley. 2007.

Visiting researcher. Cognitive Brain Research Unit. University of Helsinki, Finland. 2005.
Expert witness. 2004. Voice identification in court, assisting Dr. Peter Ladefoged.
Evaluator of the Spanish Intelligibility Test, by Dr. Daniel Kempler. Speech and Hearing Clinics, Los Angeles County/USC Medical Center.
Associate Editor. Tlalocan. 1996–97.
Associate Researcher. Indigenous Languages Seminar. Instituto de Investigaciones Filológicas, UNAM, 1994–97.
Researcher Assistant. Project: "Vocabulario trilingüe Español–Náhuatl–Otomí de Alonso Urbano (1605)". Yolanda Lastra researcher and Thomas Smith-Stark, coordinator, 'Biblioteca Novohispana de Lenguas Indígenas'. El Colegio de México. 1992–95.

Documentation efforts

Extensive field research and documentation of the phonetics and phonology of American Indian languages: Zapotec, Pima,, Northern Pame, Lowland Chontal, Kiliwa, and Yucatec Maya, among others.

Academic research and publications

Thesis

2004. "Topics in Yalálag Zapotec, with particular reference to its phonetic structures." Ph.D. dissertation. Department of Linguistics, University of California, Los Angeles.
Committee Chair: Dr. Pamela Munro
Committee: Dr. Peter Ladefoged, Dr. Edward Keenan, Dr. Matthew Gordon, Dr. Claudia Parodi

Books

Avelino, Heriberto (editor). *New Perspectives in Mayan Linguistics*. MIT Working Papers in Linguistics 59.

Selected Articles

in press. "The phonetics and phonology of nonmodal phonation variability in Yalálag Zapotec", *International Journal of American Linguistics*
in press. "Acoustic and electroglottographic analyses of nonpathological nonmodal phonation", *Journal of Voice*
under review. "México City Spanish", *Journal of the International Phonetic Association*
2009. "Intonational patterns of topic and focus constructions in Yucatec Maya", in *New Perspectives in Mayan Linguistics*. MIT Working Papers in Linguistics 59
2006. "The typology of Pamean number systems", *Linguistic Typology*, 10, 41–60
2006. Optimización dinámica y cambio morfológico en lenguas pames. [Dyanamic optimization and morphological change in Pame languages]. *Proceedings of Homenaje a Leonardo Manrique*. INAH
2003. Categorical perception of phonemic tone in Yalálag Zapotec. *International Congress of Phonetic Sciences,* Barcelona, España
2002. Sandhi tonal en Pame Norte. Análisis fonológico y evidencia fonética. *Estudios Lingüísticos sobre algunas lenguas mexicanas. Del cora al maya yucateco*, Paulette Levy (ed.). Seminario de Lenguas Indígenas, Instituto de Investigaciones Filológicas, UNAM.
1999. Morfología y Organización Prosódica en Pame Norte. *Dimensión Antropológica* Vol. 13. Instituto Nacional de Antropología e Historia
1996. El Sistema Consonántico del Pame Central. *Tercer Encuentro de Lingüística en el Noroeste*. Memorias. Serie Lingüística. Tomo 1: Lenguas Indígenas, volumen 1. Editorial UniSon.
——, Ian Maddieson and Loreta O'connor. 2009. "The phonetic structures of Lowland Chontal of Oaxaca" *International Journal of American Linguistics.*
—— and Sahyang Kim. 2004. Variability and constancy in the articulation and acoustics of Pima coronals, *Proceedings of the 29th Berkeley Linguistic Society.* University of California, Berkeley.
—— and Virgil Lewis. 2004. Pima causatives and argument structure 'saturation', *Memorias del Encuentro de Lingüística en el Noroeste. In memoriam Ken Hale*, Unison, Hermosillo, Sonora, México.

Major Documentation Project

Curriculum Vitae

First Name	Monica	Title, if any	Dr.
Family Name	Macaulay	Nationality	USA
Title of current post or study	Professor of Linguistics	Date of appointment or registration	1996
Name or employer or place of study	University of Wisconsin, Madison		

Education/Training

List your highest/latest qualification first

Dates of study	Degree	Subject	University/Institution
1982–87	Ph.D.	Linguistics	University of California, Berkeley
1980–81	M.A.	Linguistics	University of California, Berkeley
1975–79	B.A.	Linguistics (with great distinction in scholarship)	University of California, Berkeley

Employment

List your last 3 positions

Date	Position	Name of Employer
1996–present	Professor (2001–present) Associate Professor (1996–2001)	Department of Linguistics, University of Wisconsin, Madison
1988–96	Associate Professor (1994–96) Assistant Professor (1988–94)	Department of English, Purdue University
1987–88	Visiting Assistant Professor	Department of English, George Mason University

List any other significant awards, exhibitions, or other achievements relevant to this application

Relevant to expertise in Mixtec

Dissertation:

1987. *Morphology and Cliticization in Chalcatongo Mixtec.* Dissertation committee: Leanne Hinton (chair), Charles J. Fillmore, Johanna Nichols.

Books:

1996. *A Grammar of Chalcatongo Mixtec.* (Grammar with texts and dictionary; 298 pp.) University of California Publications in Linguistics, Vol. 127. Berkeley, Los Angeles: University of California Press. (revised version of thesis)

Articles and book chapters

2005 "The Syntax of Chalcatongo Mixtec: Preverbal and Postverbal." In *Verb First:On the Syntax of Verb-Initial Languages.* Andrew Carnie, Heidi Harley, and Sheila Ann Dooley (eds.), pp. 341–66. Amsterdam: John Benjamins.

1995 "The Phonology of Glottalization in Mixtec." (vith Joseph C. Salmons) *International Journal of American Linguistics* 61:38–61.

1993 "Argument Status and Constituent Structure in Chalcatongo Mixtec." *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society, Special Session on Syntactic Issues in Native American Languages*, pp. 73–85.

1990 "Negation and Mood in Mixtec: Evidence from Chalcatongo." *Anthropological Linguistics* 32:211–27.

1989 "The Plural Word in Chalcatongo Mixtec." In *Proceedings of the 25th Annual Meeting of the Chicago Linguistic Society*: 288–99.

1987 "Cliticization and the Morphosyntax of Mixtec." *International Journal of American Linguistics* 53:119–35.

1986 "Interacting Semantic Systems: Mixtec Expressions of Location." (with Claudia Brugman) In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*: 315–27.

1985 "The Semantics of 'Come', 'Go', and 'Arrive' in Otomanguean Languages." *Kansas Working Papers in Linguistics* 10(2):56–84.

1982. “Verbs of Motion and Arrival in Mixtec.” In *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*: 414–26.

Reviews

1995 Book note on *Studies in the Syntax of Mixtecan Languages, Vol. 4,* C. Henry Bradley and Barbara E. Hollenbach, eds. Summer Institute of Linguistics and the University of Texas at Arlington, 1992. Language 71:836–37.

1994 Review of *Studies in the Syntax of Mixtecan Languages, Vol. 3,* C. Henry Bradley and Barbara E. Hollenbach, eds. Summer Institute of Linguistics and the University of Texas at Arlington, 1991. Languages of the World, 8(1):68–70.

1993 Review of *Studies in the Syntax of Mixtecan Languages, Vol. 2,* C. Henry Bradley and Barbara E. Hollenbach, eds. Summer Institute of Linguistics and the University of Texas at Arlington, 1990. *Word* 44:309–11.

Fieldwork and Field Methods Courses on Mixtec and Otomanguean languages:

1997. Field methods course on Chinantec, at UW-Madison.

1992, 1985, 1982. Fieldwork on Mixtec, in Chalcatongo, Oaxaca, Mexico.

1981–87. Fieldwork on Chalcatongo Mixtec, in Berkeley, Calif.

1984. Fieldwork on Otomí, in Berkeley, Calif..

1983. Fieldwork on Zapotec, in Berkeley, Calif.

Non-Mixtec work on language documentation: grants and fieldwork on Menominee

2006–9. DEL (Documenting Endangered Languages – NEH and NSF collaboration), $310,000 for Menominee Dictionary project.

2003–6. National Science Foundation BCS-0235873, $350,000 for dictionary of Menominee.

2008, 2005, 2004, 2003, 2002, 2001, 1999, 1998. UW-Madison Graduate School Research Grants, for work on Menominee.

1998. Endangered Languages Fund, for work on Menominee.

1998–present. Fieldwork on Menominee, in Wisconsin.

Professional service

Linguistic Society of America, Executive Committee, 2006–9.

Editorial Board, International Journal of American Linguistics, 2002–present.

Declaration

I hereby confirm that:

1. To the best of my knowledge, the information provided in this application is accurate and complete.

2. I have read the relevant Information Pack, and Terms & Conditions of Award and, if a grant is made, I agree to abide by them.

3. The necessary facilities will be made available by the host institution to conduct this research.

4. I understand that it is my responsibility to request reviews of the project from the selected referees, to be returned no later than the deadline for the submission of applications.

5. If awarded the grant, I agree to develop an archiving plan in collaboration with ELAR and to deposit data and analyses developed by the project, and to assign metadata to the deposited materials.

Signature of applicant		**Date 28 June 2009**

For and on behalf of the host institution *(An appropriate person with institutional authority e.g. Head of Research or Financial Officer)*

Signature		**Date 28 June 2009**
Name (print)	**John E. Ryan**	**Institutional stamp**
Position in host institution	**Vice-Provost**	
Name of host institution	**Gettysburg College**	

Archival Holdings on Mixtec

Primary (audio) documentation of Mixtec languages (*Ethnologue* lists 45 languages in the state of Oaxaca, 5 in Guerrero, 1 in Puebla, and 1 spoken in the Oaxaca/Guerrero border) is virtually nonexistent and restricted almost entirely to elicitation and word lists in two major centers.

Archive of Indigenous Languages of Latin America (U Texas): Given the methodology of parallel elicitation of cognate lexical items across various Mixtec languages, the Josserand archive is invaluable for interlanguage documentation (dialectology) based on word lists and inflectional forms. Of the 318 individual items archived at AILLA, 164 are elicitation and 139 are wordlists; all but two items are from Oaxaca. Of the remaining material (a few items are of two types) there are 7 articles, 2 sets of fieldnotes and Josserand's dissertation. In addition there is one item indexed as a song (though it may be miscatalogued as it appears to be a discussion) and six narratives. Of the total time of 27 minutes 43 seconds for the narratives, at most half is Mixtec (the remaining is Spanish translation or discussion).

The only two items from a Guerrero Mixtec language are a Swadesh word list (9:47) and a discussion of syntaxis (approx. 100 minutes) from Tepango, municipality of Ayutla. The latter is extremely valuable and will be examined in detail in the present project by Rey Castillo.

Audio Archive of Linguistic Fieldwork, Berkeley Language Center (U California): Two major collections are stored here, both from San Miguel el Grande, Oaxaca. The first set comprises 143 recordings from 1985 (the first recording has been separated into two) totaling 34 hours, 23:21 by Leanne Hinton and Monica Macaulay. Perhaps 100 items are elicitation and the majority of the rest are discussions, mostly in Spanish. There are very few narratives and no transcriptions. The second set comprises recordings made in 1982 by Monica Macaulay. Of the 25 items (no run time is given) 21 are catalogued as "miscellaneous words and phrases".

In sum, there are almost no narrative recordings in a Mixtec language beside the 38 hours (100 items) of high quality digital recordings and time-coded transcriptions produced by Amith's and Castillo's pilot project.

Bibliography of Principal Academic Works on Mixtecan Languages
(does not include unpublished manuscripts and the "vernacular" SIL publications)
* indicates works dealing with Mixtecan languages spoken in Guerrero

Mixtec texts

Dyk, Anne. 1959. *Mixteco texts*. Norman, Okla.: Summer Institute of Linguistics. [42 texts (transcriptions and translations) from San Miguel el Grande, Oaxaca; about 20,000 words, equivalent to perhaps 4 hours of speech; no associated sound files]

Hollenbach, Barbara. 1982. A Copala Trique deluge story. *Latin American Indian Literatures* 6:114–25.

——. 1977. El origen del sol y de la luna—cuatro versiones en el trique de Copala. *Tlalocan* 7:123–70.

Hollenbach, Barbara (comp), Manuel Camilo Ramírez Santiago (narr.). 1988. *Three Trique myths of San Juan Copala*. Mexico City: Instituto Lingüístico del Verano. [3 texts (transcriptions and interlinear translations) from San Juan Copala, Oaxaca; about 3,500 words, equivalent to less than 1 hour of speech; no associated sound files]

Additional varied texts at http://www.sil.org/mexico/mixteca/00e-mixteca.htm under "Alfabetización y literatura". The texts are short and without tonal markings (e.g., "Cuento del león y el zancudo", 4 pages; "El gato y el ratoncito", 6 pages). Other materials accessible here include primers (e.g. on numbers), and miscellaneous articles (e.g. "Paradigma del verbo correr en el mixteco de Magdalena Peñasco", 10 pages).

Mixtec dictionaries

Beaty de Farris, Kathryn, et al. 2002. *Diccionario básico del mixteco de Yosondúa, Oaxaca*. Mexico City: Instituto Lingüístico del Verano. 175 pp. [about 1,000 entries, with some subentries and example phrases].

Dyk, Anna, and Betty Stoudt. 1965. *Vocabulario mixteco de San Miguel el Grande*. Mexico City: Instituto Lingüístico del Verano. 132 pp.

Good, Claude. 1978. *Diccionario triqui de Chicahuaxtla: triqui-castellano, castellano-triqui*. Mexico City: Instituto Lingüístico del Verano. 104 pp.

Pensinger, Brenda J. 1974. *Diccionario mixteco-español, español-mixteco: Mixteco del este de Jamiltepec, pueblo de Chayuco*. Mexico City: Instituto Lingüístico del Verano. 151 pp.

Stark C, Sharon (Sara), Audrey Johnson P., Benita González de Guzmán. 2006. *Diccionario básico del mixteco de Xochapa, Guerrero*. Mexico City: Instituto Lingüístico del Verano. 150 pp. [about 700 entries, with some

subentries and example phrases]
Stark Campbell, Sara, et al. 1986. *Diccionario mixteco de San Juan Colorado. Mixteco–Español / Español–Mixteco.* Mexico City: Instituto Lingüístico del Verano. 209 pp.

Mixtec grammars

SIL

Alexander, Ruth Mary. 1980. *Gramática mixteca de Atatlahuca*. Mexico City: Instituto Lingüístico del Verano. 256 pp.
Bradley, C. Henry. 1970. *A linguistic sketch of Jicaltepec Mixtec*. Norman, Okla.: Summer Institute of Linguistics. 97 pp.
Daly, John P. 1973. *A generative syntax of Peñoles Mixtec*. Norman, Okla.: Summer Institute of Linguistics. 90 pp.
*Hollenbach, Barbara E. and C. Henry Bradley, eds. 1988–92. *Studies in the Syntax of Mixtecan Languages*. 4 vols. Arlington, Tex.: Summer Institute of Linguistics and the University of Texas at Arlington. (I: Introduction; and syntactic sketches of Jamiltepec, Ocotepec, Silacayoapan; II: *Ayutla, Coatzosapan; III: *Alacatlatzala, Diuxi-Tilantongo; Concepción Pápalo Cuicatec; IV: Yosondúa, Copala Trique).

Non-SIL

Castillo García, Rey. 2007. "Descripción fonológica segmental y tonal del mixteco de Yoloxóchitl, Guerrero." Thesis para Maestría en Lingüística Indoamericana, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), Mexico City, Mexico.
Macaulay, Monica. 1996. *A grammar of Chalcatongo Mixtec*. University of California Publications. Linguistics vol. 127. Berkeley, Calif.: University of California Press. 298 pp.

Phonetics and Phonology

Aranovich, Raul. 1994. The tone system of Acatlán Mixtec and some exceptions to the OCP. *Linguistic Notes from La Jolla* 17:3–26.
Buckley, Eugene. 1991. Low-tone spreading in Chalcatongo Mixtec. In James E. Redden, ed., *Occasional Papers on Linguistics: Papers from the 1991 American Indian Languages Workshop.* Carbondale: Southern Illinois University, pp. 168–72.
Daly, John P. 1978. Notes on Diuxi Mixtec tone. *Work Papers of the Summer Institute of Linguistics* 22:98–113.
——. 1992. Phonetic interpretation of tone features in Peñoles Mixtec. Proceedings of the IRCS Workshop on Prosody in Natural Speech, August 5–12, 1992. Philadelphia, Penn.: Institute for Research in Cognitive Science, University of Pennsylvania, pp. 53–62.
Daly, John P., and Larry M. Hyman. 2007. On the representation of tone in Peñoles Mixtec. *IJAL* 73:165–207.
Gerfen, Henry James. 1996. "Topics in the phonology and phonetics of Coatzopan Mixtec. Ph.D. thesis, University of Arizona. 531 pp.
—— 2001. Nasalized fricatives in Coatzospan Mixtec. *IJAL* 67:449–66.
——, and Kirk Baker. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics* 33:311–34.
Hinton, Leanne. 1991. An accentual analysis of tone in Chalcatongo Mixtec. In James E. Redden, ed., *Papers from the 1991 American Indian Languages Workshop.* Carbondale: Southern Illinois University, pp. 173–82.
——, Gene Buckley, Marv Kramer, and Michael Meacham. 1991. Preliminary analysis of Chalcatongo Mixtec tone. In James E. Redden, ed., *Occasional Papers in Linguistics, no. 16*. Carbondale: Southern Illinois University. pp. 147–55.
Hollenbach, Barbara E. 2004. Los tonos del mixteco de Magdalena Peñasco. Available at http://www.sil.org/mexico/mixteca/magdalena-penyasco/G030-TonosMixtecoMP-xtm.htm
——. 1988. The asymmetrical distribution of tone in Copala Trique. In Harry van der Hulst and Norval Smith, eds., *Autosegmental Studies on Pitch Accent*. Dordrecht: Foris, pp. 167–82.
——. 1987. La duración vocálica en el trique de Copala: Un análisis abstracto. *SIL-Mexico Workpapers* 8:15–29.
——. 1985. Vowel length in Copala Trique: An abstract laryngeal analysis. *IJAL* 51:455–57.
——. 1984. "The phonology and morphology of tone and laryngeals in Copala Trique." Ph.D. thesis, University of Arizona.
——. 1974. Reduplication and anomalous rule ordering in Copala Trique. *IJAL* 40:176–81.
Hunter, Georgia G., and Eunice V. Pike. 1969. The phonology and tone sandhi of Molinos Mixtec. *Linguistics* 47:24–40.
Iverson, Gregory K., and Joseph C. Salmons. Mixtec prenasalization as hypervoicing. *IJAL* 62:165–75.

Macaulay, Monica, and Joseph C. Salmons. 1995. The phonology of glottalization in Mixtec. *IJAL* 61:38–61.
Mak, Cornelia. 1958. The tonal system of a third Mixtec dialect. *IJAL* 24:61–70.
——. 1953. A comparison of two Mixtec tonemic systems. *IJAL* 19:85–100.
——. 1950. A unique tone perturbation in Mixteco. *IJAL* 16:82–86.
Martlett, Stephen A. 1992. Nasalization in Mixtec languages. *IJAL* 58:425–35.
Meacham, Michael. The phonetics of tone in Chalcatongo Mixtec. In James E. Redden, ed., *Papers from the 1991 American Indian Languages Workshop.* Carbondale: Southern Illinois University, pp. 156–67.
North, Joanne and Jäna Shields. 1977. Silacayoapan Mixtec phonology. In William R. Merrifield, ed., *Studies in Otomanguean Phonology. SIL Publications in Linguistics* 54:21–33.
*Overholt, Edward. 1961. The tonemic system of Guerrero Mixteco. In Benjamin F. Elson and Juan Comas, eds., A William Cameron Townsend. Mexico City: Centro de Investigaciones Antropológicas de México. pp. 597–626.
*Pankratz, Leo, and Eunice V. Pike. 1967. Phonology and morphotonemics of Ayutla Mixtec. *IJAL* 33:287–99.
Pike, Eunice V., and John H. Cowan. 1967. Huajapan Mixtec phonology and morphophonemics. *Anthropological Linguistics* 9(5):1–15.
Pike, Eunice V., and Thomas Ibach. 1978. The phonology of the Mixtepec dialect of Mixtec. In Mohammad Ali Jazayery, Edgar C. Polomé, and Werner Winter, eds., *Linguistic and Literacy Studies in Honor of Archibald A. Hill, Vol 2: Descriptive Linguistics*, pp. 271–85. The Hague: Mouton.
Pike, Eunice V., and Joy Oram. 1976. Stress and tone in the phonology of Diuxi Mixtec. *Phonetica* 33:321–33.
Pike, Eunice V., and Priscilla Small. 1974. Downstepping terrace tone in Coatzospan Mixtec. In Ruth Brend, ed., *Advances in Tagmemics*. Amsterdam: North-Holland. pp. 105–34.
Pike, Eunice V., and Kent Wistrand. 1974. Step-up terrace tone in Acatlán Mixtec. In Ruth Brend, ed., *Advances in Tagmemics*. Amsterdam: North-Holland. pp. 81–104.
Pike, Kenneth L. 1944. Analysis of a Mixteco Text. *IJAL* 10: 113–38.
——. 1945a Tone puns in Mixteco. *IJAL* 11:129–39.
——. 1946a. Another Mixteco tone pun. *IJAL* 12:22–24.
——. 1946b. 'The Flea': Melody types and perturbations in a Mixteco song. *Tlalocan* 2:128–33.
*Zylstra, Carol F. 1980. Phonology and morphophonemics of Mixtec of Alacatlazala, Guerrero. *SIL Mexico Workpapers* 4:15–42.

Morphosyntax

Anderson, Lynn. 1993. You can say that again: Repetition in Alacatlatzala Mixtec. *SIL Mexico Workpapers* 10:38–53.
de León, Lourdes. 1986. A dialectal view of Mixtec noun classifiers: Productivity and fossilization. In Schott DeLancey and Russell S. Tomlin, eds., Proceedings of the Second Annual Meeting of the Pacific Linguistics Conference. Eugene, Ore.: Department of Linguistics, University of Oregon. pp. 337–61.
Eberhardt, Roy. 1999 “Questions and inversion in Ocotepec Mixtec.” *Workpapers of the Summer Institute of Linguistics, University of North Dakota Session*. 43:1–20.
*Hills, Robert A., and William R. Merrifield. 1974. Ayutla Mixtec, just in case. *IJAL* 40:283–91.
Hinton, Leanne. 1982. How to cause in Mixtec. In Monica Macaulay et al., eds., *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*, pp. 354–63.
Hollenbach, Barbara E. 1997. Covert transitive-intransitive verb pairs in Copala Trique. SIL-Mexico Workpapers 12:1–10.
——. 1995a. Semantic and syntactic extensions of body-part terms in Mixtecan: The case of ‘face’ and ‘foot’. *IJAL* 61:168–90.
——. 1995b. A preliminary catalog of focus devices in Mixtecan languages. *SIL Mexico Workpapers* 11:1–16.
——. 1992. Parsing relative clauses in Copala Trique. In Shin Ja J. Hwang and William R. Merrifield, eds., *Language in Context: Essays for Robert E. Longacre*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington, pp. 537–51.
——. 1990. Semantic and syntactic extensions of Copala Trique body-part nouns. In Beatriz Garza Cuarón and Paulette Levy, eds., *Homenaje a Jorge A. Suárez*. Mexico City: El Colegio de México, pp. 275–96.
——. 1984. Reflexives and reciprocals in Copala Trique. *IJAL* 50:272–91.
——. 1976. Tense-negation interplay in Copala Trique. *IJAL* 42:126–32.
Macaulay, Monica. 2005. “The Syntax of Chalcatongo Mixtec: Preverbal and Postverbal.” In *Verb First:On the Syntax of Verb-Initial Languages.* Andrew Carnie, Heidi Harley, and Sheila Ann Dooley (eds.), pp. 341–66. Amsterdam: John Benjamins.
——. 1993. Argument status and constituent structure in Chalcatongo Mixtec. *Proceedings of the Nineteenth Annual*

Meeting of the Berkeley Linguistics Society, Special Session on Syntactic Issues in Native American Languages: 73–85.
—— 1990. Negation and mood in Mixtec: Evidence from Chalcatongo. *Anthropological Linguistics* 32(3):211–27.
—— 1989. The plural word in Chalcatongo Mixtec. Proceedings of the 25th Annual Meeting of the Chicago Linguistic Society: 288–99.
—— 1987. Cliticization and the morphosyntax of Mixtec. *IJAL* 53:119–35.
—— 1985. On the semantics of 'come', 'go' and 'arrive' in Otomanguean languages. *Studies in Native American Languages IV. Kansas Working Papers in Linguistics* 10(2):56–84.
—— 1982. Verbs of motion and arrival in Mixtec. In Monica Macaulay et al., eds., *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*: 414–26.
——, and Claudia Brugman. 1986 "Interacting Semantic Systems: Mixtec Expressions of Location." *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*: 315–27.
Kuiper, Albertha, and William R. Merrifield. 1975. Diuxi Mixtec verbs of motion and arrival. *IJAL* 41:32–45.
Kuiper, Albertha, and Velma B. Pickett. 1974. Personal pronouns in Diuxi Mixtec. SIL Mexico Workpapers 1:53–58.
Merrifield, William R., and Betty J. Stoudt. 1967. Molinos Mixtec clause structure. *Linguistics* 32:58–78.
Pensinger, Brenda, and Larry Lyman. 1975. Some Eastern Jamiltepec Mixtec phrase constructions. *IJAL* 41:158–61.
Pike, Kenneth L. Analysis of a Mixteco text. *IJAL* 10:113–38.
Williams, John L. 1996. Tezoatlán Mixtec motion and arrival verbs. *IJAL* 62:289–305.
——. 1993a. Four conjunctions in Tezoatlán Mixtec. *SIL-Mexico Workpapers* 10:68–84.
——. 1993b. The fronting of noun and adverb phrases in Mixtec of Tezoatlán. *SIL-Mexico Workpapers* 10:85–111.

Historical

Dürr, Michael. 1987. A preliminary reconstruction of the Proto-Mixtec tonal systems. *Indiana* 9:189–206.
Josserand, Judy Kathryn. 1982. Mixtec Dialect History. Ph.D. thesis, Tulane University. 711 pp.
Hollenbach, Barbara. 1977. Phonetic vs. phonemic correspondence in two Trique dialects. In William R. Merrifield, ed., *Studies in Otomanguean Phonology*. Dallas, Texas: Summer Institute of Linguistics, pp. 35–67.
Longacre, Robert E. 1957. Proto-Mixtecan. *Indiana University Research Center in Anthropology, Folklore and Linguistics, Memoir* 15. Bloomington, Ind.: Indiana University.
Mak, Cornelia, and Robert Longacre. 1960. Proto-Mixtec phonology *IJAL* 26:23–40.

Principal Material on Guerrero Mixtec

Alacatlatzala: Anderson 1993; Zylstra 1980, 1991 ("A syntactic sketch of Alacatlatzala Mixtec," in Hollenbach and Bradley, III:1–177); 1 short SIL vernacular publication listed at *Ethnologue*
Alcozauca: Stark, Johnson, and González de Guzmán 2006; 3 short SIL vernacular publications listed at *Ethnologue*
Ayutla: Hills 1990 ("A syntactic sketch of Ayutla Mixtec," in Hollenbach and Bradley, II:1–260), Hills and Merrifield 1974, Pankratz and Pike 1967; 11 short SIL vernacular publications listed at *Ethnologue*
Metlatonoc: Overholt 1961; no vernacular publications listed at *Ethnologue*
Yoloxóchitl: No academic or vernacular publications besides the master's thesis of Rey Castillo (2007)

Additional works cited in the proposal

Campbell, Lyle, and Martha C. Muntzel. 1989 "The structural consequences of language death." In Nancy C. Dorian, ed., *Investigating Obsolescence: Studies in Language Contraction and Death.* Cambridge: Cambridge University Press, 181–96.
Carnie, Andrew, Heidi Harley, and Sheila Ann Dooley, eds. 2005. *Verb First: On the Syntax of Verb-Initial Languages*. Amsterdam: Johns Benjamins.
Chávez-Peón, Mario E. 2008. Phonetic cues to stress in a tonal language: Prosodic prominence in San Lucas Quiaviní Zapotec. *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*.
Dryer, Matthew S. n.d. "Word order". To appear in Timothy Shopen, ed., *Clause Structure: Language Typology and Syntactic Description, vol. 1*. Cambridge University Press.
Fishman, Joshua. 1991. "How threatened is 'threatened'? (A typology of disadvantaged languages and ameliorative priorities)." In Joshua Fishman, *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages.* Philadelphia: Multilingual Matters, 81–121.
Smith Stark, Thomás. 1995. "El estado actual de los estudios de las lenguas mixtecanas y zapotecanas." In Leonardo Manrique, Yolanda Lastra and Doris Bartholomew, eds., *Panorama de los estudios de las lenguas indígenas de México*, 2 vols., Colección Biblioteca Abya-Yala núm. 16. Quito, Ecuador: Ediciones Abya-Yala,

pp. 5–186.
Suárez, Jorge A. 1983. *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.