This course will examine the principles and practices of the field of inquiry referred to as “Documentary Linguistics”, using data that I have collected and experiences that I have had while working on several languages in Yobe State, Nigeria. Students will use information from this course to do their own language documentation projects.

Linguistics has become a broad discipline. Not only are there subfields such as phonetics, phonology, morphology, syntax, semantics and “hyphenated” subdisciplines, such as socio-linguistics and psycho-linguistics, there are different methodological approaches. Among these are “formal” or “theoretical” linguistics (the focus being on language as a formal system to be described using rules and similar formalisms—the general approach of the UCLA Linguistics Department), “functional” or “discourse” linguistics (the focus being on the properties of discourse in human interaction and communication—the general approach of the UCLA Applied Linguistics Department), and, emerging over the past decade and a half, “documentary linguistics”. Nikolaus Himmelmann (Gippert, Himmelmann & Mosel 2006:1) says that “a language documentation is a lasting, multipurpose record of a language”. The methodology of documentary linguistics seeks to assemble relatively copious records, typically as audio and/or video recordings, of a language in as wide a range of its uses as possible (conversation, narrative, speeches, ritual language, etc.). Focus of nearly all work in documentary linguistics is on “endangered” languages, though the principles and practices could just as well be applied to widely spoken languages. Documentary linguistics does not concentrate on detailed analysis of data, but rather on “annotation” that will make the data interpretable and accessible by subsequent users. These may be the documenters themselves, other people with a variety of research interests (linguists, anthropologists, folklorists), or people of the community in which the language is spoken. The field of documentary linguistics pays considerable attention to the types of media used for collecting data, archiving of (annotated) data (hence the “lasting record” in the quote above), and the methodology of collecting “naturalistic” data (as opposed to direct elicitation or more indirect means of collecting data, such a completing questionnaires, with the intention of analyzing particular linguistic structures or functions). Another tenet of documentary linguistics is to have speakers of the target language be part of the data collection and processing activity.

Since 2000, I have been working in Yobe State, Nigeria on a group of six to eight languages (depending on whether one calls some of them “dialects” of one language or closely related “languages”). The languages are Bade (two dialects), Bole, Duwai,
Karekare, Ngamo (two dialects), and Ngizim. They all belong to the Chadic family, of which Hausa is by far the largest and best-known member. The Yobe State Chadic languages represent two sub-branches of Chadic that are only distantly related to each other.

I will confess that I had never heard of *documentary linguistics* as a defined field of linguistic inquiry until the last year or two, yet it turns out that, in my Yobe State projects, much of what I have been doing is language documentation pretty much as defined by the proponents of this field. I have made extensive use of video and audio recording, I have assembled data of varied types (ethnography, folktales, songs, naturalistic narratives based on pictures, ethnographical and ecological vocabulary, etc.), and all the work has involved educated native speakers as true collaborators in data collection, analysis, and processing. I have issues with some of the philosophical and methodological concepts of the “official” tenets of documentary linguistics, and I will discuss these issues during the course, but my work closely fits with the broad picture of this field. This course will use the data and experiences of my work on the Yobe languages as a point of departure for introducing what is now a dynamic field of inquiry in linguistics.

**Course Work**

Class meetings will comprise two 1:50 lectures per week. Course work and grading are as follows:

- Four written assignments: 30%
- Final exam: 20%
- Course project and class presentation: 50% (about 40% and 10% respectively)

Written assignments will be handed out about every two weeks (due the following week) and will total 30% of the grade. They will be approximately equally weighted. See the Week by Week Course Outline for tentative distribution dates and topics for assignments. The final exam will be a mix of essay and data questions.

The course project will be a Language Documentation project on a language chosen by the student, in consultation with the instructor:

**Language:** Ideally the language chosen will be a language or variety of a language with which the student has some prior familiarity but does not speak fluently. The two most likely candidates for the language would be (1) a language that the student has studied in school as a foreign language or (2) a heritage language that the student does not speak with adult native-like fluency. Many UCLA students come from families where a heritage language is spoken, and often within such families there will be relatives or family friends who speak closely related languages or dialects that the student does not speak well but is familiar with. This would an ideal choice for the course project.

**Documentation:** The student will make an audio or video recording of a native speaker narrating a text. Length of the text can vary, but somewhere between 1000-2000 words would seem appropriate. Topics for the narration are unlimited, though it should be appropriate to the language, culture, and age of the speaker. Typical topics would be a traditional folk-tale, a personal narrative, an ethnographic narrative about customs, or recounting a story in a culturally appropriate cartoon strip or video. The documentation will consist of the following elements:

**Meta data:** The field of documentary linguistics emphasizes inclusion of complete meta data with a documentation. This includes information on the language (where spoken, classification, dialect, etc.), the speaker(s) (age, sex, education, special knowledge, etc.), circumstances of data collection (location, public performance or
one-on-one with the documenter, etc.), equipment used, storage method (format of digitization, text editor used, etc.), and perhaps other information that could be relevant to future users of the data.

**Transcription:** The student will transcribe the text (with the help of the narrator if need be) using a Romanized transcription system. If the language has a standard Roman orthography, that is fine, but an explanation of the phonetic values of the symbols should be part of the documentation. If the language does not have a writing system at all or uses a non-Roman orthography, the transcription should be phonemic, using IPA or American structuralist symbols or some adaptation thereof.

**Translation and interlinear glossing:** The transcribed text will be accompanied by both free English translation and interlinear morpheme-by-morpheme glossing. It is usually convenient to have the full text and free translation on facing pages and the text repeated separately with interlinear glossing.

**Lexicon:** The student will prepare a lexicon of approximately 50 words (more if the student chooses) from the text with lexical entries arranged and containing the types of information that will be discussed around Week 6. The words chosen for the lexicon should be representative of word classes relevant to the language.

**Morphology, syntax:** A couple of aspects of morphology and/or syntax that are represented in the text should be chosen and expanded. For example, if the text contains past tense verb forms, the student could use those as a basis for giving a full paradigm of relevant past tense verbs. If the text contains a question-word question (“Who is there?”, “What are you doing?”), the student could use these as a basis for describing the syntax of question-word questions.

**Information on content and/or context:** The documentation will include something about how the type of text and its content relates to the culture where the language is spoken.

**Special issues:** Any special problems that came up should be mentioned, for example, word segmentation (lexically separate words may coalesce in speech), code-switching to non-native words which the speaker wanted to expunge in transcription, false starts and other breaks in the narrative, etc.

**Prospectus:** By Thursday of Week 5, students should have e-mailed Prof. Schuh one-paragraph descriptions of their projects, including the language, the speaker to be used, the type of medium for recording (audio cassette, digital audio, video), and an idea of the nature of the text.

**Class presentations:** During Weeks 9-10, students will give a presentation of 10 minutes or so of what they are doing in their documentation. Ideally, there should be some “show-and-tell”, with a recording of the text and a few words about interesting discoveries.

**Alternative projects:** Students should make every effort to create a documentation of the type outlined here. If this proves impossible for some reason, an alternative would be to create a documentation of data that was collected in the Yobe Languages Project, in consultation with Prof. Schuh. Prof. Schuh would also be willing to listen to other proposals, but they must make use of the documentation techniques that will be discussed in the course.

**THE WRITTEN DOCUMENTATION AND RECORDING WILL BE DUE AT THE TIME OF THE FINAL EXAMINATION.**
**Week by Week Course Outline**

Since this is a new course, this outline is a tentative order of events. Some topics may take more time than allotted, some may take less. It may also turn out that the natural flow of events will call for reordering or restructuring of some topics. Nonetheless, I hope we get to all of them some way or another.

“GHM” = Gippert, Himmelmann, & Mosel (2006)

**WEEK 1**
- GHM Chapter 1 (not required)

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
</table>
| Tu, March 30 | Overview of the course  
Language *Documentation* and Language *Description*               |
| Th, April 1  | African languages, Chadic languages, and the languages of Yobe State, Nigeria |
|             | **Assignment 1** Looking at some examples of Language Documentation on the web |

**WEEK 2**
- Russell G. Schuh, Proposal for NSF Grant BCS-0553222
- Larry M. Hyman, “Field work as a state of mind” (Newman & Ratliff 2001:Chapter 1)
- GHM Chapters 2-3 (not required)

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, April 6</td>
<td>The Yobe Languages Research Project [YLRP]</td>
</tr>
<tr>
<td>Th, April 8</td>
<td>Field work</td>
</tr>
</tbody>
</table>

**WEEK 3**
- GHM Chapter 11

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, April 13</td>
<td>Getting started: lexicon and transcription</td>
</tr>
<tr>
<td>Th, April 15</td>
<td>Getting started: looking at a text</td>
</tr>
<tr>
<td></td>
<td><strong>Assignment 2</strong> Collection and presentation of a short text</td>
</tr>
</tbody>
</table>

**WEEK 4**
- GHM Chapters 4, 9, 13
- Shobhana L. Chelliah, “The Role of text collection and elicitation in linguistic fieldwork” (Newman & Ratliff 2001:Chapter 7)

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, April 20</td>
<td>Data and data organization in language documentation</td>
</tr>
<tr>
<td>Th, April 22</td>
<td>Texts: types of texts, collecting texts</td>
</tr>
</tbody>
</table>

**WEEK 5**
- GHM Chapter 8

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, April 27</td>
<td>Texts: transcribing and annotating texts</td>
</tr>
<tr>
<td>Th, April 29</td>
<td>Texts: what can texts be used for</td>
</tr>
<tr>
<td></td>
<td><strong>Prospectus</strong> E-mail to Prof. Schuh outlining plan for class project</td>
</tr>
</tbody>
</table>
WEEK 6

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, May 4</td>
<td>Lexicon: expanding lexicon; lexical entries</td>
</tr>
<tr>
<td>Th, May 6</td>
<td>Lexicon: lexical entries (cont.), output of a lexicon</td>
</tr>
</tbody>
</table>

**Assignment 3** Creating a small lexicon

WEEK 7

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, May 11</td>
<td>Elicited data: phonology</td>
</tr>
<tr>
<td>Th, May 13</td>
<td>Elicited data: morphology</td>
</tr>
</tbody>
</table>

WEEK 8

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, May 18</td>
<td>Elicited data: syntax</td>
</tr>
<tr>
<td>Th, May 20</td>
<td>Presentation and distribution of data descriptions</td>
</tr>
</tbody>
</table>

**Assignment 4** Formal description of a selected data set

During Weeks 9-10, if there is time, we may return to texts, discussing issues such as analyzing folktales in terms of motifs, difficulties in understanding ethnographic texts by an "outsider" (the linguist fieldworker!), and/or content and musical settings of songs.

WEEK 9

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, May 25</td>
<td>Student presentations</td>
</tr>
<tr>
<td>Th, May 27</td>
<td>Student presentations</td>
</tr>
</tbody>
</table>

WEEK 10

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu, June 1</td>
<td>Student presentations</td>
</tr>
<tr>
<td>Th, June 3</td>
<td>Student presentations; wrap up; course evaluations</td>
</tr>
</tbody>
</table>

References


The site itself doesn’t have much information, but it has links to concepts and organizations concerned with Language Documentation/Documentary Linguistics.


A useful summary document about the field of documentary linguistics.

[http://titus.uni-frankfurt.de/ld/](http://titus.uni-frankfurt.de/ld/)

The official website accompanying Gippert et al. (2006).

[http://www.humnet.ucla.edu/humnet/aflang/Yobe/yobe.html](http://www.humnet.ucla.edu/humnet/aflang/Yobe/yobe.html)

The website for the Yobe Languages Research Project. Material from the Yobe Languages Research Project will be the illustrative material for this course.
SOME ORGANIZATIONS INVOLVED IN DOCUMENTING LANGUAGES

DELAMAN: Digital Endangered Languages and Musics Archives Network
  http://www.delaman.org/

DoBeS: Dokumentation Bedrohter Sprachen (Documentation of Endangered Languages)
  http://www.mpi.nl/DOBES/

EAGLES: Expert Advisory Group on Language Engineering Standards
  http://www.ilc.cnr.it/EAGLES96/home.html

ELAP: Endangered Languages Academic Programme
  see HRELP

ELAR: Endangered Languages Archive Programme
  see HRELP

ELDP: Endangered Languages Documentation Programme
  see HRELP

E-MELD: Electronic Metrastructure for Endangered Languages Data
  http://emeld.org/

HRELP: Hans Rausing Endangered Language Project
  see also ELDP, ELAP, ELAR
  http://www.hrelp.org/

IMDI: ISLE Meta Data Initiative
  http://corpus1.mpi.nl/ds/imdi_browser/

ISLE: International Standards for Language Engineering
  http://corpus1.mpi.nl/ds/imdi_browser/

OLAC: Open Language Archives Community
  http://www.language-archives.org/

Some other relevant acronyms

HLT: Human Language Technology
  see http://www.ilc.cnr.it/EAGLES/isle/ISLE_Home_Page.htm

NIMM: Natural Interaction and Multimodality
  see http://www.ilc.cnr.it/EAGLES/isle/ISLE_Home_Page.htm
SOME WEBSITES LINKING TO ONLINE OR DOWNLOADABLE LANGUAGE DOCUMENTATION (dictionaries, grammars, etc.)

http://www.rosettaproject.org/ (go to the “Internet Archive” link)
Lots of scanned documents of various types that are in the public domain.

http://www.yourdictionary.com/languages.html
Links to many online and downloadable dictionaries.

http://www.yourdictionary.com/grammars.html
Links to online and downloadable grammars. Many—maybe most or all—of the links are to teaching grammars rather than descriptive grammars.

http://corpus1.mpi.nl/ds/imdi_browser/ (= IMDI link above)
A confusing and annoying site to use, but it brings together links to the majority of the “acronym” sites above that claim to do “language documentation” in the documentary linguistics sense. This site is part of the Max Planck Institute for Psycholinguistics, which maintains the DoBeS (Documenting Endangered Languages) database. You have to register with the MPI to use the DoBeS links, and even that registration does not give access to lots of the stuff on the DoBeS site. You probably have to register separately for most, if not all the other sites that are linked to this one website. Frankly, just exploring this site, I found very few documents that were (immediately) usable. You can view and/or download some VERY big raw video and audio files that have no transcription or other documentation that I could find other than the metadata. Maybe I just have to learn how to use this site, but it makes me wonder about the language documentation enterprise (1) if access is so complicated and restricted and (2) if the data itself doesn’t have any obvious use other than to see/listen to a language you don’t know anything about.