

CÁTEDRA JOAQUIM HELIODORO DA CUNHA RIVARA  
GOA UNIVERSITY

**PRO 126 - METHODS OF LANGUAGE DOCUMENTATION**  
(1<sup>ST</sup> FEBRUARY TO 14<sup>TH</sup> FEBRUARY 2018)

**HUGO C. CARDOSO**  
*Universidade de Lisboa*  
hugocardoso@gmail.com

**COURSE DESCRIPTION**

Language documentation refers to the collection and treatment of language samples, be it for purposes of the preservation of knowledge, specific research agendas, the development of educational resources, or many others. Documentation is very often linked to linguistic description – the production of descriptive works such as grammars or dictionaries –, of which it is often the necessary first step. Currently, with a growing conscience of the rapid decline in the world's linguistic diversity and of the fluidity in the linguistic repertoire of many communities, as well as the rapid development of the Digital Humanities, the impetus for documentation and archiving has become more urgent.

This course will equip students with the essential tools to carry out linguistic documentation that is both coherent and useful, while at the same time safeguarding the integrity of the participants' intellectual property and responding to other ethical concerns. It will introduce tools and develop skills which are useful for research in Linguistics but also for the study of other disciplines of the Humanities and Social Sciences, including Anthropology, History, Sociology, Education, and Musicology.

## **COURSE SYLLABUS**

PRO 126 - METHODS OF LANGUAGE DOCUMENTATION

1<sup>nd</sup> February

Class 1

### **INTRODUCTION TO LANGUAGE DOCUMENTATION**

This class introduces the history and objectives of linguistic documentation and the constitution of digital linguistic corpora, and the reasons behind the current focus on this research practice. Topics covered include:

- Linguistic documentation in the 19<sup>th</sup> century and now;
- The contraction of the world's linguistic diversity;
- Language documentation and language preservation;
- Specific programmes and opportunities for documentation.

---

2<sup>th</sup> February

Class 2

### **ETHICAL ASPECTS OF DOCUMENTATION**

Any research dealing with human knowledge and practice needs to ensure the willingness of the participants to collaborate and to have their contributions made accessible to others. Therefore, it is essential to obtain informed consent from all participants. In this class, we will discuss the ethical issues that arise when doing linguistic documentation, the ethical limitations imposed on data collection and dissemination, as well as the need to ensure that the community's interests and needs are combined with the researcher's own interests.

---

5<sup>th</sup> February

Class 3

### **LINGUISTIC FIELDWORK**

In most cases, language documentation presupposes the conduction of fieldwork, often in locations and among communities very distant from those of

the researcher. While the success of fieldwork depends, to a large extent, on engagement, common sense and sharp observation, there are certain techniques and skills that can be trained in order to ensure proper and effective data collection. Topics covered include:

- The Observer's Paradox;
- Various fieldwork tasks;
- Interview and elicitation sessions;
- Collaborative work and training;
- Orthography development;
- Elicitation techniques.

---

7<sup>th</sup> February

Class 4  
**AUDIO AND VIDEO RECORDING**

In this class, we will focus on the equipment used for audio and video recording of language and learn the basic techniques to ensure good quality results in a fieldwork setting. Students will have a chance to practice in the classroom. Issues covered include:

- What to record;
- Appropriate equipment;
- Maintenance of equipment;
- File formats;
- Resolutions, light and noise.

---

9<sup>th</sup> February

Class 5  
**DATA-COLLECTION PRACTICE**

This is a practical class in which students will apply the technical knowledge they acquired in the previous sessions to a real-world situation. The class will be conducted outside the classroom and students are expected to approach informants and obtain audio and/or video recordings of conversations. The resulting materials will then be discussed in class, to identify possible difficulties and solutions to them.

10<sup>th</sup> February

Public lecture  
**A DIGITAL ARCHIVE OF SRI LANKA PORTUGUESE**

This public lecture reports on a major documentation project, currently ongoing, aimed at the documentation of Sri Lanka Portuguese, a Portuguese-lexified creole language spoken primarily in the Eastern Province of Sri Lanka by the Portuguese Burghers. The project ‘Documentation of Sri Lanka Portuguese’, developed at the University of Lisbon (Portugal) in cooperation with Sri Lankan institutions and various collaborators around the world, and funded by the ‘Endangered Languages Documentation Programme’ (SOAS, United Kingdom), is concerned not only with the collection of language samples but also with recording several aspects of Portuguese Burgher culture and social practice. We will introduce the language and the community and explain the documentation activities which have already been implemented to constitute a coherent, online-accessible archive: interviews, elicitation, recordings of song and dance performances, sociolinguistic surveys, and the development of orthographic conventions.

---

12<sup>th</sup> February

Class 6  
**PROCESSING LANGUAGE DATA**

In this class, we will learn how to process the recordings made in the field. We will train how to copy the audio and/or video files to the computer, how to transform them, and how to prepare them for transcription and annotation. We will introduce common software used in linguistic documentation, including ELAN and FLEx.

---

14<sup>th</sup> February

Class 7  
**PROCESSING EXERCISE**

In this class, students will use ELAN and FLEx to produce a short annotated/glossed linguistic corpus, using either their own recordings or others provided by the instructor.

## CLASS MATERIALS

Readings – a selection of chapters (provided by the instructor) from:

- Chelliah, Shobhana & William de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.
- Gippert, Jost, Nikolaus Himmelmann & Ulrike Mosel. 2006. *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.
- Thieberger, Nicholas. 2012. *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press.

Software:

- Audacity: <http://www.audacityteam.org/download/>
  - Transcriber: <http://trans.sourceforge.net/en/presentation.php>
  - FLEx: <http://fieldworks.sil.org/flex/>
  - HandBrake: <https://handbrake.fr/downloads.php>
  - Avidemux: <http://fixounet.free.fr/avidemux/download.html>
  - ELAN: <https://tla.mpi.nl/tools/tla-tools/elan/>
- 

## ASSESSMENT

- Active participation in class\* 20%
- Annotated corpus\*\* 40%
- Research proposal\*\*\* 40%

\* Includes doing the assigned readings and the preparatory work and being prepared to discuss their contents in class.

\*\* Produced during the classes.

\*\*\* A research proposal to conduct the documentation of a specific endangered language, in which all the technical and ethical aspects discussed are properly addressed.