# USER GUIDE TO ESDS DATA DEPOSIT

Project title: **Social networks, target language interaction and second language acquisition during the year abroad: a longitudinal study**

ESRC award number: RES-062-23-2996

Principal Investigator/ contact for inquiries: Professor Rosamond Mitchell, Modern Languages, University of Southampton, Southampton SO17 1BJ, UK. Email: [R.F.Mitchell@soton.ac.uk](mailto:R.F.Mitchell@soton.ac.uk)

Subject area: Linguistics

## GENERAL DESCRIPTION OF DATA

These data comprise a corpus of advanced learner French and advanced learner Spanish and will be of interest to researchers into second language acquisition for these two languages.

The participants were Anglophone university students majoring in languages at a UK university (29 for French, 27 for Spanish). The participants completed a battery of assessments on 6 occasions over a period of 18 months, before, during and after spending a compulsory academic year abroad, in France, Spain or Mexico. The data collection schedule was as follows:

|  |  |
| --- | --- |
| Pre-test (Letter code a) | May 2011 |
| Visit 1 abroad (Letter code b) | November 2011 |
| Visit 2 abroad (Letter code c) | February 2012 |
| Visit 3 abroad (Letter code d) | June 2012 |
| Post-test 1 (Letter code e) | October 2012 |
| Post-test 2 (Letter code f) | February 2013 |

The data being made available to other researchers comprises learner outputs in French or Spanish, collected on all 6 occasions, using a set of 3 tasks:

1. Oral interview

Participants took part in a semi-structured interview in French or Spanish (Task O) on each occasion, with a member of the research team. The interviews varied in detail on each occasion but all centred on participants’ hopes, plans and actual experiences of residence abroad. (Interview schedules are available by request to the research team.)

1. Picture-based narrative

Participants told a short story in the target language, based on a sequence of pictures, and designed to promote production of past tense forms. Three parallel stories were developed: Cat Story ( Task C), Sisters Story (Task S) and Brothers Story (Task B). ( The picture sequences can be obtained on request to the research team.)

1. Argumentative essay

Participants wrote a timed essay in French or Spanish in response to a predetermined prompt, and were asked to produce 200 words over a 15 minute period. The essay was written on computer with no access to dictionaries or spell checker (but access to accented characters needed for French or Spanish). Three versions were developed:

“Do you think that gay couples have the right to marry and adopt children?” (Task G)

“Do you think that marijuana should be legalised?” (Task D)

“To encourage healthy eating, higher taxes should be imposed on soft drinks and junk food” (Task F)

Administration of these tasks was scheduled as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Pre-Test (a) | Oral Interview (Task O) | Cat Story (Task C) | Writing (Task G) |
| Visit 1 (b) | Oral Interview (Task O) | Sisters Story (Task S) | Writing (Task D) |
| Visit 2 (c) | Oral Interview (Task O) | Brothers Story (Task B) | Writing (Task F) |
| Visit 3 (d) | Oral Interview (Task O) | Cat Story (Task C) | Writing (Task G) |
| Post-Test 1 (e) | Oral Interview (Task O) | Sisters Story (Task S) | Writing (Task D) |
| Post-test 2 (f) | Oral Interview (Task O) | Brothers Story (Task B) | Writing (Task F) |

For each participant 18 task performances are available from each learner participant. In addition, 10 native speakers of French and 10 native speakers of Spanish each carried out all tasks on a single occasion (letter code n), for purposes of comparison.

Due to space limitations however the corpus currently deposited with ESDS comprises **the Learner Pretest data ONLY** for both French and Spanish. Further data is available via the project website at <http://langsnap.soton.ac.uk> .

## FILE FORMATS

Performances of the oral interview (Task O) and the picture narratives (Tasks C, S and B) were audiorecorded in .wav format using Roland R-09HR digital recorders. The audio files were then converted into MP3 format and these versions have been contributed to ESDS.

All oral interviews and picture narratives were also transcribed by members of the research team, using the CHAT format developed by the CHILDES project (conventions are accessible at <http://childes.psy.cmu.edu> ). The resulting .cha files have been contributed to ESDS. In order to read these files it is necessary to download/ instal the free CLAN program from the CHILDES website.

The .cha files of the oral interviews and picture narratives were tagged for parts of speech using the MOR and POST programs developed by the CHILDES project and associates (Christophe Parisse is acknowledged as developer of the MOR program for French).These tagged files have also been contributed to ESDS. Again to read these tagged files it is necessary to download/ instal CLAN.

For each oral task performance, therefore, the database includes 3 file types: .cha, .mor.pst.cex, and .mp3 .

The text files of the argumentative essays (Tasks G, D and F) were converted to .cha format and tagged similarly for parts of speech using MOR and POST. For each written performance, the database includes 2 file types: .cha and .mor.pst.cex .

## FILE HEADERS

Each file has a unique header detailing the task (identified by a single upper case letter code), the participant (identified by a 3-digit code), the occasion (identified by a single lower case letter code) and (for the oral production tasks) the researcher administering the task (2-4 upper case letter code). The file type is indicated by the file extension (.cha, .mor.pst.cex, .mp3 ).

French learner participants are identified by 3-digit codes from 100 to 129; Spanish learner participants by 3-digit codes from 150-180; and native speaker participants by 3-digit codes from 130-139 (French) and from 185-194 (Spanish).

For example:

Administered at Post-test 1

Transcript file in .cha format

Oral interview task

O115eKMcM.cha

Researcher initials

Participant 115 (French)

Tagged file in .mor.pst.cex format

Administered to native speaker

Writing

task D

D191n.mor.pst

Native speaker participant 191 (Spanish)

## FILE ORGANISATION

The files available through ESDS are as follows:

French Pretest Interview MP3

French Pretest Interview Tagged

French Pretest Interview Transcripts

French Pretest Narrative MP3

French Pretest Narrative Tagged

French Pretest Narrative Transcripts

French Pretest Writing Tagged

French Pretest Writing Transcripts

Spanish Pretest Interview MP3

Spanish Pretest Interview Tagged

Spanish Pretest Interview Transcripts

Spanish Pretest Narrative MP3

Spanish Pretest Narrative Tagged

Spanish Pretest Narrative Transcripts

Spanish Pretest Writing Tagged

Spanish Pretest Writing Transcripts

## ANONYMISATION

Care has been taken to anonymise the data generously contributed by project participants. The guidelines used to anonymise soundfiles and transcripts are attached as an Annex to this document. The researchers would be grateful if users could bring any anonymisation errors to their attention.

## ACKNOWLEDGEMENTS

All users of this material are requested to make an explicit acknowledgement of the source in any resulting publication, by reference to the project website: <http://langsnap.soton.ac.uk> .

Any use of CLAN programs to analyse the material should also be acknowledged, citing:

MacWhinney, B 2000 *The CHILDES Project: Tools for analysing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

(fuller guidance is given at <http://talkbank.org/share/rules.html> )

Ros Mitchell [R.F.Mitchell@soton.ac.uk](mailto:R.F.Mitchell@soton.ac.uk)

Nicole Tracy-Ventura [nkt@usf.edu](mailto:nkt@usf.edu)

Kevin McManus [Kevin.mcmanus@york.ac.uk](mailto:Kevin.mcmanus@york.ac.uk)

29.01.14

# Anonymization (LANGSNAP Project)

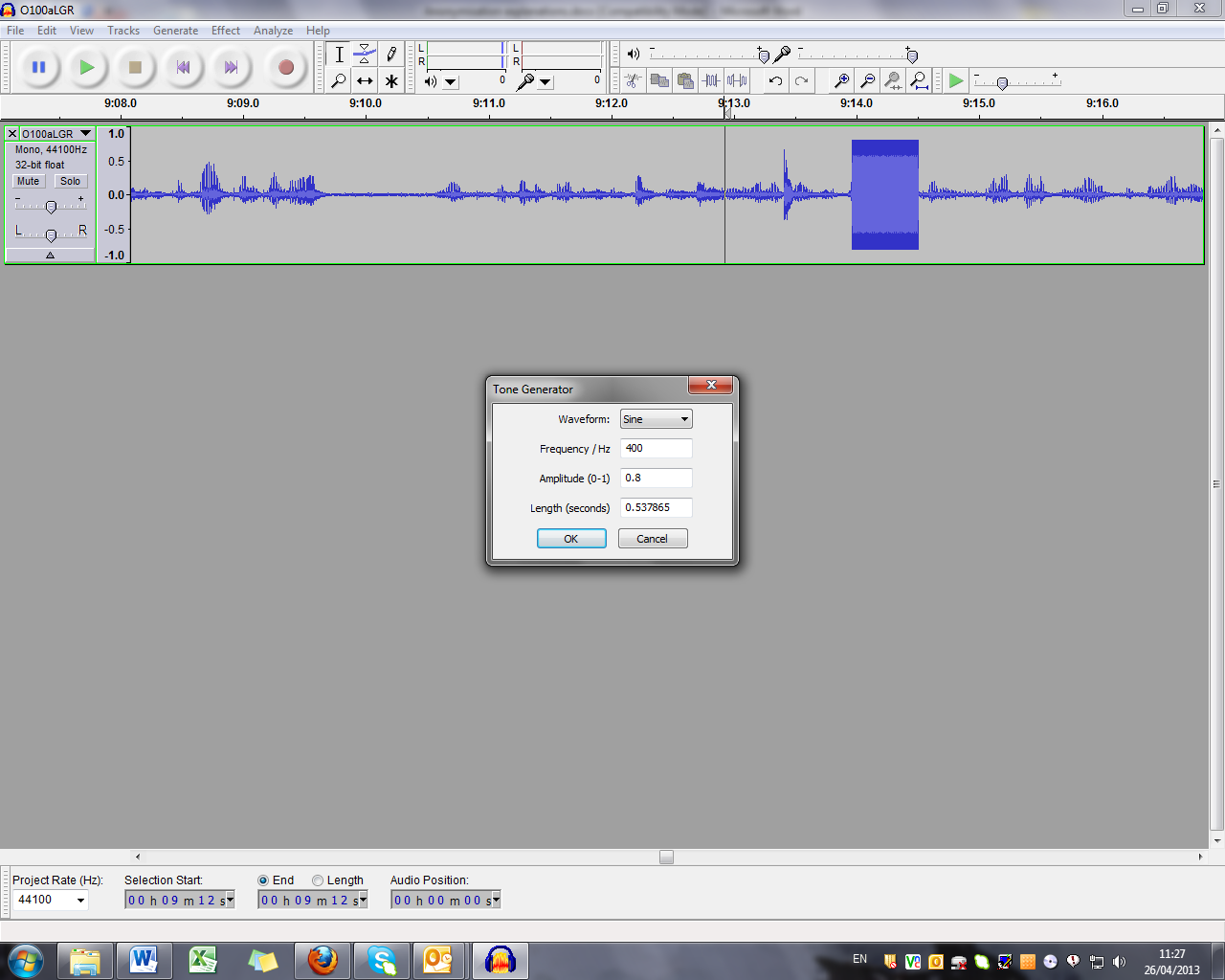
Audio files (.mp3 and .wav) are anonymized by removing name and place information. Name and place information is removed from the audio and replaced by a tone (300Hz) using Audacity. It is important not to delete the original audio file. In the transcripts, place and name information is removed and replaced with the labels *ville* (for place) and *nom* (for name), respectively[[1]](#footnote-1). The files audio files to be anonymized are: Spoken Narratives, Oral Interviews and Writing.

## Process for audio anonymization

* You will anonymize the .wav only and then convert it to mp3 using Audacity
* Take the transcript that has been marked up for anonymization. Find its audio equivalent in .wav format. Make a copy of the original .wav and save it here in the appropriate subfolder: J:\HUMS Research\YA Project\Anonymization
* You will work only on the copy and not the original.
* Open the audio to be anonymized (.wav) with Audacity. Using the marked transcript, you will find the text that needs removing and replace it with a 300Hz Tone. The text that needs replacing is highlighted in yellow.

|  |  |  |
| --- | --- | --- |
| **The identified** | **Transcript mark-up** | **Audio mark-up** |
| Name (any name) | Nom | Tone 300Hz |
| Place (any place) | Ville |

* To anonymize the audio, you locate the sounds in the file and highlight the excerpt on the spectogram that needs removing. You then go to ‘generate’ in the menu bar and select ‘tone’. Select ‘Sine’ for waveform and enter 400 for Frequency. When you click ‘ok’ the selected sound will be replaced by a tone.



When you have replaced all identified words with a tone, you save the file by using the export function (located in ‘file’ on the menu bar). Export the file as WAV.

## Guidelines for anonymizing transcripts

Below are a set of agreed conventions for removing personal information from the transcripts so that participants are not identifiable.

|  |  |
| --- | --- |
| **What you find in the text** | **How you anonymize it** |
| **For people**  Female name (e.g. Nicole)  Male name (e.g. Kevin)  Male/Female name, or unsure of sex | Name\_F  Name\_M  Name\_? |
| **For places**  Institution (e.g. Université de Perpignan)  Street name (e.g. Highfield Road)  City/Town/Village (e.g. Nancy)  Southampton  University of Southampton  Not to be anonymized:  Regions (e.g. Hampshire)  Countries (e.g. Italy)  Capital cities (e.g. Paris) | Institution  Street  City  Home City  Home Institution |

RM/NTV/KMcM 15 May 2013

1. Or Spanish equivalents [↑](#footnote-ref-1)