

**Written assignment:
Bilingual language acquisition --
Analyzing data from the CHILDES database**

Go to the CHILDES webpage (Child language data exchange system)
<http://childes.psy.cmu.edu/>

From the CHILDES page, please **download** the following files and programs:

- The CLAN program to read and analyze transcripts
- CLAN manual, explaining how to use the program (pdf file)
- CHAT manual, explaining the transcription format (pdf file)
- transcripts of a bilingual Spanish-English-speaking child, collected by Margaret Deuchar (click on "Database," then "Downloadable transcripts," then "Bilinguals," and finally, "Deuchar-zipped.zip")
- the "Bilingual Corpora" guide to the CHILDES data on bilingual language acquisition, which contains a description of Deuchar's dataset; on the starting page, click on "Database Manuals," and then pick "Bilingual Acquisition."

Deuchar's data are from a longitudinal study of a Spanish-English bilingual child in England, recorded between the ages of 1 year, 3 months and 3 years, 3 months. The data set is divided into two subsets, entitled "English" and "Spanish," based on the languages used by the adults. After downloading the data, I recommend that you copy all the transcript files (".cha" files) into a single directory, such as the "lib" folder that is set up when you install CLAN. This will facilitate searching all the files at once. In order for the CLAN program to search through the files, the directory containing the files has to be chosen as working directory and lib directory in the command window.

Using the program and the database, please perform the following analysis steps and write up your findings in a brief paper, which should also address the questions stated below.

Using the "commands" window in CLAN, identify the Mean Length of Utterance (MLU) for the child at three different stages for interactions in each language. (Use the `mlu` command, following the instructions in the CLAN manual). Note that you can search several files at once by using * as a placeholder for other letters in the file name. For example, if you enter `freq +t*CHI 86*.cha` or `mlu +t*CHI 86*.cha`, it will search through all utterances by the child in all files with names beginning with "86" (here all transcripts from recordings made in 1986). If you state a specific file name, the program will search only that file. So, for example, if you enter `freq +t*CHI 860928sf.cha` you will get a list of all the words uttered by the child during the one recording session included in this file.

Identify three referents for which the child knows the word in both English and Spanish (e.g. *dog* and *perro* ‘dog’), and indicate at what age they are being used. Are there any such word pairs in the data from the one-word stage?

To identify word pairs, it makes sense to first generate lists of English words uttered by the child, and then search for these words in the translation tier (%mor) of the Spanish-language files.

For example, the following search (using the *kwal* command)

```
kwal +s"*book*" +t*CHI +t%mor *s*.cha
```

will yield all utterances in which the English word *book* is found either in utterances of the child (+t*CHI), or in English translations of the child’s utterances (on the %mor tier which contains a morpheme-by-morpheme interlinear translation of Spanish utterances), searching only the Spanish-language files (i.e. files that contain the letter *s* in the file name *s*.cha). The * symbol indicates that the search string ignores adjacent letters, so it will find the word *book* whether it is by itself or whether there is some other coding attached to it (e.g. n|libro&masc=book). The string +s"book" would only turn up instances where book is in isolation, i.e. where the letter sequence is between two empty spaces.

Identify three examples in which the child uses English words with her father, as well as three examples where she uses Spanish words in interactions with her mother and grandmother (who address her in English). You can do this by using the *freq* command to give you wordlists of the individual files. How do the adults react to her use of the “wrong” language? (To read the transcript, open the file in CLAN).

Finally, answer the following questions in at least one paragraph each, giving evidence for your analysis:

- Drawing on the readings that you read on bilingual language acquisition, how would you characterize the child’s acquisition process and her parents’ language socialization strategy? (compare DeHouwer 2007)
- Do you think that the child is dominant in English or Spanish, or is her language acquisition balanced between the two languages?
- Do the data provide evidence for the Unitary Language System Hypothesis (see Genesee 2001: 155), or is there evidence for Early Differentiation (see Meisel 2001)?

In formatting your paper, please follow the handout on formatting guidelines.