



**Computer Simulation of Bilingual Infants Learning Catalan and Spanish**  
**Sanaa Ounis (COL 2018)**  
**Advisor: Yakov Kronrod**

This summer I worked as a research assistant under the mentorship of Dr. Yakov Kronrod on the project, “Computer Simulation of Bilingual Infants Learning Catalan and Spanish.” There is a surprising learning behavior for bilingual infants learning Catalan and Spanish. Infants will perceive language depending on their linguistic input. Spanish-Catalan bilingual infants will temporarily be unable to distinguish between two Catalan phonemes, in a habituation task, but will succeed in an anticipatory eye movement (AEM) task. I used a lexical distributional model with different types of input to show how infants can fuse these two phonetic categories. Afterward, I worked a computer simulation of the habituation and AEM tasks to show the discrepancy in infants’ discrimination of these sounds.

Spanish-Catalan bilingual adults tend to make speech errors when they conflate the two phonemes together, and the children of these adults will learn these mistakes. I ran simulations with the errors that these bilingual adults make in order to exhibit how children mistake these two vowels. I also ran simulations with Catalan input and Spanish input to show that these learning mistakes come from errors that bilingual adults make. After running this learning simulation, I ran a program that simulates the actual child behavior in laboratory tests. The program included a simulation of the habituation and AEM tasks.

The habituation task is when tokens of either phoneme category are played continuously until the infant becomes dishabituated. The AEM task is when infants hear nonsense words containing either phoneme and would watch Elmo appear behind an occluder. The side in which Elmo appeared corresponds to the side the sound came from. Their surprisal and eye movement is measured when a sound is played with a delayed appearance of Elmo.

In order to model infant data, I had to use a fully Bayesian updating model for the prior. The prior consisted of what the “infant” would already know in the real world during training and then what the “infant” learned from training during testing. I also systemically explored different parameters and inputs to give optimal results to model this phenomenon.

Throughout my research experience this summer, I learned more about computational linguistics. I didn't know much about this field of study before starting this project. I am now more comfortable with coding using various computer program languages. As a statistics minor, I was glad I could analyze the data using many statistical and probabilistic models. Without this opportunity I would not have been able to further develop my skills, hence I am very thankful for PURM.