Documentation of R scripts:

**LME –**

Model: Linear Mixed Effects Model

Data: same\_probes.csv

This is derived from the loose dataset. In the loose dataset, the same topic is shared, but the participants are different.

Includes all participants in levels 3,4, and 5 that speak either Spanish, Korean, or Chinese. All participants who only participants in level 3,4, or 5 and do not progress through the levels are included as well as those who did progress through the levels.

There is one regular model and another model that accounts for interaction between the level\_id and language.

Anon\_id and question\_id are treated as random factors.

**ANOVA-Test –**

Model: ANOVA two-way

Data: whole

This is derived from the whole dataset. Only Chinese, Korean, and Spanish speakers are included, and only those from levels 3,4, and 5 are included.

**LMER-SCAloose –**

Model: Linear Mixed Effects Model

Data: SCAloose

This data set derives from the loose dataset. The participants were put into three groups (3-4,4-5,3-5). Participants could be in multiple groups. If a participant only had entries in level 3, they were not put into any groups. From there, the essays were ran through L2SCA.R to find different syntactic measures.

Each of the syntactic measures calculated from L2SCA.R was used as a y variable for a model.

This script was used to analyze each level grouping (3-4,4-5,3-5).