Julio Cesar Lopez Otero

Final project – peer review

**Peer review 1:**

**Verb Frequency and DOM in Heritage Speakers of Spanish**

**Esther Hur**

Good abstract, I understand the topic you are examining and the importance it has in the heritage language domain. However, it would be nice to state a research question and a hypothesis to guide the reader through the study.

Good methods section, you discuss your participants’ background information, as well as the tasks that you used. I would be interested in knowing the background of these participants (or their parents). Is there any dialectal variation in the use of differential object marking in Spanish? Also, I am curious as to why you set the cut between intermediate and advanced speakers in 40 out of 50 points in the DELE. Maybe it would have been interesting to include proficiency as a continuous variable. That would have helped with the unbalanced number of participants in the two proficiency groups. Regarding the tasks, you mention an Acceptability Judgment Task but don’t report it in the results section.

Interesting results. Finding lexical frequency effects in the production of syntactic structures is very meaningful to the research on heritage language. Also, great plot. I liked the use of the jitter points. However, the plot does not have a title and I do not understand what the x axis represents. What is ‘countLog’? Why does it range from 7 to 9?

Your discussion is short but interesting. I see that it appears in a different format than the rest of the paper.

**Questions:**

I see the chunk of code below in your R script. What is the ‘log’ function and what does it do? Why did you use it?

ept\_elog <- ept\_df %>%

group\_by(., participant, proficiency, countfreq, groupCon, response, item) %>%

summarize(., n = 8,

wDOM = sum(response),

woDOM = n - wDOM,

eLog = log((wDOM + 0.5) / (n - wDOM + 0.5)),

wts = 1 / (wDOM + 0.5) + 1 / (n - wDOM + 0.5),

countLog = log(countfreq))

You only found frequency effects in the intermediate group and I see in your R script that all verbs tested were different. Maybe they only used DOM in frequent verbs because they consider the ‘a’ as a preposition that makes part of verb (cuidar a, encontrar a), as a sort of lexical chunk similar to ‘depend on’?

--

**Comments for Prof. Casillas**

I am not able to run Esther’s R script on my computer. Maybe the function here() is not working on my computer.

I wonder why Esther reports her results like this: “The model provides main effects of group (F (2) = 206.6 , p < 0.001), but no main effect of frequency. However, there were group by frequency interaction (F (1) = 3.48, p < 0.1).” instead of using the pattern we have used in the past (beta = X, SE = X, z = X, p = X). What is the difference between those? Are both valid in the journals to which we usually submit our work?

Also, the paper shows the code used to generate the plot above the plot. She should have used echo = FALSE to hide it.

I also wonder why the format of the discussion section looks different, like part of a chunk code. I checked her .Rmd and did not find any particular code that would make it look like that.

My main concern with this paper is not strictly related to statistics, but with lexical frequency. Intermediate heritage speakers might produce DOM as part of a ‘verb lexical chunk’ that they have acquired regardless of the animacy and definiteness features involved in the Spanish DOM. Reporting the results of the AJT can limit the strength of these critics. However, in further studies it would be interesting to test the same verbs in two different contexts: (1) required DOM and (2) ungrammatical DOM (La abuela encuentra a la nieta vs La abuela encuentra la aguja). I believe this could solve this issue.

**Peer review 2:**

**Acquisition of null subjects**

**Michele Goldin**

Very interesting and informative abstract and introduction sections. I also liked that you clearly stated your research questions and hypotheses. That helped me as a reader to understand the goal of your study.

Great methods section. It was very thoughtful to exclude participants exposed to Caribbean varieties.

Good results section. Everything is organized in subcategories that respond to each research question. Great plots. They are very clear: titles, labeled x and y axes, and explanations. On your second plot, your paper includes a warning derived from your .Rmd code. I wonder what could be done to avoid that. The last plot is very informative.

Your discussion is also very interesting. Morphology is an indicator for syntax in English, but that correlation is not clear in Spanish because Spanish null/overt subjects are ruled by pragmatic constraints. I believe you could argue that morphology is an indicator for syntax in both languages (as they accept both null and overt subjects in Spanish, the two grammatical options), but not for pragmatics in Spanish. That is consistent with the distinction between ungrammatical and infelicitous that you present above.

**Questions**:

You follow two different patterns to report your glm() results. First, you use (Beta = X, SE = X, z = X, p = X), and at the end you report them like this: “In Spanish, there were no main effects and no interaction. In English, however, there was a main effect of group (F(2) = 12.81, p < .002) and a main effect of proficiency and group (F(1) = 5.18, p < 0.02) as well as a marginally significant group by English proficiency interaction (F(2) = 5.28; p = .071).” Why do you report them differently? I thought the models were all tested with the same code and gave the same kind of results.

I have a question concerning your fillers. If the question used to trigger the participants’ response to the acceptability judgment task was ‘Who said it better?’, how did that work with questions aiming meaning instead of form? Were the participants expected to choose the puppet that did not lie as the person who said it better? Is it appropriate to mix meaning vs form in acceptability judgment tasks? Also, were the participants’ responses as expected in the fillers? In other words, were they sensitive to the lies of the puppets?

--

**Comments for Prof. Casillas**

Michele reported the results from glm() in two different ways:

“The monolingual group responded with an overt pronoun significantly more often than the L2 group (β = 0.91; SE = 0.32; z = 2.83; p < .005) and significantly more often than the heritage group (β = -0.98; SE = 0.30; z = -3.27; p < .002).” (p. 8)

“In English, however, there was a main effect of group (F(2) = 12.81, p < .002) and a main effect of proficiency and group (F(1) = 5.18, p < 0.02) as well as a marginally significant group by English proficiency interaction (F(2) = 5.28; p = .071).” (p. 11)

I see in her R script that the results above come from:

Call:

glm(formula = cbind(overt, null) ~ group, family = "binomial",

data = df\_counts\_wide[df\_counts\_wide$language == "eng", ])

Deviance Residuals:

Min 1Q Median 3Q Max

-2.9400 -0.7186 0.0000 0.6505 2.7423

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.58779 0.19720 2.981 0.00288 \*\*

groupmo 0.91043 0.32138 2.833 0.00461 \*\*

grouphl -0.07696 0.25604 -0.301 0.76373

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 87.242 on 46 degrees of freedom

Residual deviance: 74.435 on 44 degrees of freedom

AIC: 176.94

Number of Fisher Scoring iterations: 4

While the results below come from:

> anova(mod\_null, mod\_group, mod\_prof, mod\_add, mod\_full, test = 'Chisq')

Analysis of Deviance Table

Model 1: respFactor ~ 1

Model 2: respFactor ~ group

Model 3: respFactor ~ profEn

Model 4: respFactor ~ profEn + group

Model 5: respFactor ~ profEn + group + profEn:group

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 375 469.39

2 373 456.59 2 12.8069 0.001656 \*\*

3 374 459.42 -1 -2.8340 0.092290 .

4 372 451.41 2 8.0113 0.018213 \*

5 370 446.13 2 5.2832 0.071249 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Why are they tested and reported differently? I thought that the anova function helped us determine the most suitable model and that was what we needed to report. Maybe because the first paragraph reports on overall model results while the second focuses on main effects?

What can be done to avoid that warning message above the second plot?

It would be interesting to look at the progression of morphology and null/overt subjects across time by making the age variable continuous. Instead of using years, she could use months of age.

Finally, it would be interesting to include any kind of production data, even if it is spontaneous. It would be interesting to compare production with comprehension and test whether your current results are consistent with the participants’ production.