Review of the contribution *Control Stimuli in CS*

Overall, the contribution addresses the important question of methodology issues in research. In particular, acceptability judgements in code-switching data are analyzed here, claiming that a baseline comparison of acceptability is needed in order to detect variation.

I believe the paper needs to complement and discuss some important issues in the study of code-switching, as I will address below. The test setting is constructed very neatly and seems to be very accurate. However, the main limit of the study is the composition of the participant group, which is way too heterogeneous in my view, so that the results lose credibility.

The paper thus needs major revisions.

**Context-related comments**

Page 2, line 76 ff: Do you only consider studies on English/Spanish? Otherwise, for adjective placement, you might also refer to Cantone & MacSwan (2009)


Page 4, examples: Why don’t you consider d. His friends have estudiado espanol? Is it linked to your discussion in Page 7? See below

Page 5, second paragraph: Did you consider that participants might consider something different as unacceptable than the switching point you address and thus judge some different linguistic phenomenon than the one you mean?

Page 5, section 3.1: In my view, 20 participants with so many differences is a way too heterogeneous group. I consider this a great weakness of the study.

Page 6, section 3.2: Did you ask whether participants use to code-switch themselves? What is the concrete definition of acceptable here? It sounds strange, it is possible? I believe all these different explanations are very confusing for the participants.

Page 7, examples 3a and b: Are you saying these examples MUST be ungrammatical or are you saying they are not possible? Is this in line with MacSwan’s model? Line 297: please cite studies on this debate.

Page 8, lines 306-307: Again, please quote WHO considers these switches as unacceptable.

Page 8, results: Why aren’t there any analyses with respect to the language profile? Please analyze whether the groups you make differ w.r.t. proficiency in the languages, age of acquisition, self-rated proficiency and exposure. These aspects could matter a lot!

**Comments on the format/typos:**

Page 1, line 41: delete second ‘that’

Restructure References alphabetically