Review of “Control stimuli in experimental code-switching research”

Summary:
This paper is a methods paper that expands on previous research by González-Vilbazo et al. (2013) on methodological aspects of code-switching (CS) research, especially with the use of acceptability judgment tasks (AJTs) to examine CS speech. The paper focuses, in particular, on control stimuli, that is, the stimuli used to compare a structure whose (un)grammaticality is under investigation. The paper has three main contributions:

1. The research design includes three Spanish-English CS that are present in natural-speech corpora and that are largely regarded as grammatical in previous research (complex-sentence switches, subject-predicate switches, and direct-object switches) and two Spanish-English CS that are largely considered ungrammatical in previous research (pronoun switches, present perfect switches) for an initial analysis to establish if these two large categories (grammatical vs ungrammatical switches) are significantly different from each other and whether differences are found within each category (complex-sentences switches vs. subject-predicate switches). With data from 20 Spanish heritage speakers, the paper reports significant differences between grammatical and ungrammatical stimuli but no differences within grammatical or within ungrammatical stimuli. Establishing these two comparison points is an important contribution to test the grammaticality of a switch for three reasons: (i) research in the past has only compared to a grammatical sentence, thus, possibly missing a mid-tier category, (ii) some switches do not have a natural comparison, and (iii) having a standard of grammatical vs. ungrammatical sentences across studies can facilitate comparisons across studies.

2. Additionally, this design shows that the three types of grammatical switches and the two types of ungrammatical switches were treated similarly within each category and, therefore, they can and should be used as control stimuli to have more variety in the stimuli.

3. The paper examines individual results and uses the control stimuli to group participants into those that distinguish between grammatical and ungrammatical stimuli and those that do not. Although they do not explore further why some participants do not distinguish, this analysis provides the basis for exclusion of participants who do not respond to the format of testing of AJTs.

Broad comments:

Background and motivation: This methods paper contributes to the literature in CS in a significant way. It is well-placed in the literature as a continuation of the methods paper by González-Vilbazo et al. (2013). In particular, it identifies a gap on the development of stimuli in AJTs in CS research. The gap is clearly identified (a need for methodological solutions to address heterogeneity among the participant results and/or how to operationalize (un)acceptability) and the paper addresses the gap successfully. Nonetheless, a few references are needed to support some of the claims. In particular, in page 7, lines 297-299, the sentence can be further supported by reference to work on auxiliary verb switches more generally and switching have/haber in Spanish-English CS. In the lines that follow, claims about present-progressive switches also need references added (page 8).
**Methodology:** The paper is methodologically sound, as it tests the relevant conditions for establishing clearly grammatical and clearly ungrammatical sentences and the benefits for having these. The design is conscientious. Using real sentences from the Bangor Miami corpus, for instance, is a strength in the design. There are few methodological issues that can be addressed as limitations, as they are minor.

- For the ungrammatical sentences, it would be useful to include that these were not present in the Bangor Miami corpus, to make it comparable to the grammatical sentences, which were extracted from this corpus. The author can examine the corpus and, if there are no examples, indicate this in the description of the materials. If there are examples, a brief explanation can be included in a footnote.
- For the pronoun stimuli, since the materials were presented visually, how did the author control for prosodic stress? A clarifying footnote should be enough to explain how they assured that the pronouns lacked the defining characteristics of strong pronouns (specially for Spanish pronouns).
- Is there any reason why or any concern about all filler stimuli being subject-predicate switches?
- The authors used a 7-point scale and did not include a ‘not sure’ option. Usually with uneven-numbered scales it is a good idea to have a ‘not sure’ option so that they use the middle option to indicate that it is somewhere in between completely acceptable and completely unacceptable and not that they are not sure how to rate the sentence.
- There is no item analysis and a low number of items per condition (at least considering other subdisciplines within linguistics) so there may be some variability that has not been examined. For the present-progressive switches, for instance, for the distinguishing group, could some items have received higher ratings than other and the results given be the result of certain items being rated high and others low? This can fall within the purview of future research recommendations.
- Related to this idea, too, is the use of ANOVAs instead of mixed-effects models, which have become the standard in the field. The type of analysis performed in mixed-effects models addresses directly differences across individuals. A lot of the problems derived from heterogeneity can be addressed statistically with mixed-effect models but, in this case, since it is the focus of the methods paper it does not seem necessary. In the case of variation between items, though, since it is not directly examined in the paper, the author may consider using a mixed-effects model with item as a random factor. Alternatively, this issue can just be noted as a limitation.
- Similarly, the number of participants and the heterogeneity of the group (ages and places of origin) can also be briefly acknowledged in the discussion of limitations.

**Discussion and directions for future research:** With respect to present-progressive switches, more discussion could be added as to how to proceed when the three groups of participants are identified: those who accept, those who reject, and those in between. This is especially important in light of the discussion about acceptability being categorical. Additionally, some of the methodological aspects mentioned in the previous section could be discussed in this section, especially those related to limitations and directions for future research.

Overall, the paper addresses the three aims it is set out to explore:
- Explore the differences between grammatical and ungrammatical stimuli
- Participant heterogeneity and exclusion of participants who do not use the scale
- Establish a baseline comparison of (un)grammaticality that allows for a middle tier because, if no ungrammatical items had been included, the middle group would have resembled the group that rated them as ungrammatical

Although there are some minor methodological recommendations, the paper offers important design suggestions that other researchers are likely to adopt.

**Specific comments:**
Text should be checked for typos. I only found a couple:
- On page 2, line 78 there is a double parenthesis `)))`
- On page 11, line 419 the second ‘unacceptable’ should be ‘acceptable’ instead.