

Session 5: „*Making a deal with the stats devil*”

Task 1:

A group of researchers conduct an eye-tracking study on the processing of subject- vs. object-relative clauses in prototypical native speakers of Turkish raised and living in Turkey (N=60) versus heritage speakers of Turkish from the United Kingdom (N=60). In the review process, the reviewers severely criticize the fact that the heritage group contains a substantial number of people (N=23) with an age of arrival in the host country of 11 or higher. They argue that the term “*heritage speaker*”, according to established definitions in the field (e.g. Rothman, 2009; Benmamoun, Montrul, & Polinsky, 2010), only applies to individuals who arrived in the host country in early childhood. As a result, your heritage group consists of a complicated mixture of heritage speakers of Turkish and native speakers of Turkish who grew up in Turkey and then emigrated to the UK in early adulthood (or even later), which makes the results from the heritage group near-uninterpretable. They suggest doing the study again with a heritage sample which consists exclusively of participants who arrived in the host country in early childhood.

- 1) How would you respond to the above criticism in a response letter? Would you follow the suggestion and re-do the study accordingly, or can you come up with other possibilities to address this concern (such as specific additional statistical analyses)?
- 2) One of the reviewers is not satisfied with your response letter, and argues that it is impossible to draw conclusions about apples from a sample which consists of a complicated mixture of apples and oranges. She thus suggests that you should apply a clear cut-off criterion for age of acquisition in the host country, and exclude all participants who are not heritage speakers from your heritage sample. She adds that,

given that your sample is relatively large, a clean subsample consisting only of real heritage speakers would still be large enough. Do you agree with this rationale, or can you see any problems with it? Could such an approach also be potentially detrimental for the field of heritage-language processing?

Task 2:

You review a study investigating the production of affixed inflected verb forms in prototypical native speakers (N=19), native French L2 learners (N=18), and heritage speakers of Kinyarwanda living in France and Belgium (N=15). The authors report the results from a cloze task consisting of 50 items, in which the participants had to fill in the correct inflectional affixes. The results from this task are shown in Table 1 below.

Table 1. Proportion of correct items (and SDs in brackets) in the cloze task by participant group.

	L1 control	L1 heritage	L2
Accuracy (out of 1.00)	.86 (.14)	.73 (.25)	.68 (.27)

To avoid the problem of variance being related to the mean in proportion data, the authors analysed the results using logit mixed-effects models. The analyses show a significant difference between the L1 control and L1 heritage groups ($p < .05$) and between the L1 control and L2 groups ($p < .05$), but not between the L1 heritage and L2 groups ($p = .11$).¹ The authors thus suggest that the L1 heritage group patterns with the L2 group, with both of these groups being outperformed by the L1 control group. They conclude that, at least with regard to inflectional morphology, the amount of exposure to the target language is relatively more important than age of acquisition onset.

¹Please note that, in a real journal article, the results from the logit mixed model would of course have to be written up in a considerably more detailed way, with at least a full table reporting the results for the fixed effects contained in the model, as well as the model formula, in detail. I have shortened this section for teaching purposes, with only the p-values reported. For the sake of this exercise, please assume that the way in which the logit-mixed model analysis was conducted is technically correct.

- 3) As a reviewer, what would you criticize about the study described above? What would your overall recommendation be, and why?