
How to handle variation in linguistic judgment data

Gert-Jan Schoenmakers¹ & Roeland van Hout²

¹*Universiteit Utrecht*, ²*Radboud Universiteit*

g.t.schoenmakers@uu.nl, roeland.vanhout@ru.nl

When collecting linguistic judgments from larger groups of participants and for largesets of items, the dataset is bound to be ‘contaminated’ by individual variation: people may respond to stimuli differently and test sentences (in the same condition) may elicit different responses. Various researchers have shared recommendations to maximally reduce such variation (e.g. Häussler & Juzek 2021; Christianson et al. 2022). However, there will always be variation in language data. An alternative perspective is therefore to ‘handle’ variation and include it as much as possible in our analyses to investigate if the observed patterns go beyond noise, and, if so, which factors may explain them.

In this talk, we discuss which steps to take in dealing with variation in linguistic judgment data. We argue that the treatment of datasets in their full form, including all variation, is a fundamental component of the scientific enterprise. The talk consists of four main topics:

- i) choices in the treatment of judgment data (data type, z-transformation),
- ii) choices in random-effects structure design and methods to investigate random structures as relevant sources of information,
- iii) explorations of the variation patterns (visualizing the distribution of the effects, performing cluster analyses), and
- iv) performing reliability analyses on item sets distributed according to within-item counterbalanced designs (which is impossible with the tools of classical test theory, but can be done with generalizability theory).

The reliability analyses are an important contribution of the talk, as they provide a way to chart to what extent datasets are generalizable and replicable. These analyses allow us to assess whether participants are consistent in their judgments – and similarly for item sets. Moreover, we can calculate the numbers of participants and items required for good reliability in future studies, based on variation patterns of existing studies.

The data we use to illustrate our main points come from a two-level dataset and various datasets from experiments using 2×2 designs (see additional pages). Based on these data, we show that variation should be treated as an integral part of the dataset.

References: Häussler, J. & T. Juzek (2021). Variation in participants and stimuli in acceptability experiments. In G. Goodall (ed.), *The Cambridge handbook of experimental syntax*. CUP, 97–117. • Christianson, K. et al. (2022). What if they’re just not that into you (or your experiment)? *Psychology of Learning and Motivation* 76, 51–88.