We are interested in determining the robustness of a particular kind of semantic priming effect by considering different data-processing pathways that are arguably equivalent, rather than selecting one particular pathway. We are asking your input to determine which of the pathways are suitable to include.

*Research question*

As in the first survey, we will again focus on semantic priming. In general, people are faster to recognize a target (e.g., *dog*), when it is preceded by a related prime (e.g., *cat*) compared to an unrelated prime (e.g., *car*). It is often assumed that the magnitude of the priming effect varies depending on how strongly the prime (*cat* in the above example) and target (*dog* in the example) are related. For instance, *cat*-*dog* may be a more strongly related pair compared to *finger*-*toe*. In this study, we seek to examine whether such *item-level* priming effects are stable across languages. More specifically, if items exhibit a strong priming effect in English do they also exhibit a strong priming effect in German, and vice versa for items yielding weak priming effects? We will only focus on priming effects in terms of response time, not accuracy.

Critically, there are different ways to process such data. We now want to examine whether the potential stability of *item-level priming* across languages, or lack thereof, is *robust* across different equivalent data-processing pathways. Next, we will briefly summarize the procedure of the original data collection, which is important to understand the different potential data-processing options that were identified.

*Procedure original data collection*

We will rely on data from a recent (ongoing) study by Buchanan et al. (2022) which is currently investigating semantic priming across 10+ languages using equivalent, translated stimuli. Participants (adults) had to perform a so-called continuous lexical decision task. On each trial, participants saw a letter string, which either formed an existing word or a nonword. Participants needed to decide as quickly and accurately as possible whether the letter string was an existing word by pressing either *Z* or */* on a QWERTY keyboard (or similar pattern on the native language keyboard). When no response was provided within 3 seconds, the trial was automatically terminated. Participants got 10 practice trials followed by a total of 800 test trials, split up in blocks of 100, using an intertrial interval of 500 ms. After each block, participants could take a break. There were 400-word trials and 400-nonword trials. 150-word trials involved a critical target (e.g., *dog*), half of which were preceded by a related prime trial (e.g., *cat*), and the other half by an unrelated prime trial (e.g., *car*). The other trials were fillers. Participants saw a particular stimulus only once during the study, and whether a given target was preceded by its related or unrelated prime was determined at random. If you require additional information, feel free to contact us, or you can also consult Buchanan et al.’s paper here: <https://osf.io/q4fjy/>

*Goal of the current study*

In the first part of the survey, we will ask you to evaluate different choices pertaining to *data-processing with a particular analysis in mind*. In other words, we will introduce the final steps of the analysis pathway, and we ask you to evaluate different potential data-processing pathways. More specifically, response times to the critical targets will be z-transformed for each participant separately (i.e., every participant’s arithmetic mean response time to critical targets will be subtracted from their response time at each target trial and the result will be divided by the participant’s standard deviation again only using critical trials). Next, we will separate related and unrelated trials for each target, after which we subtract their arithmetic mean z-transformed response times (aggregated across participants), for example: . This step will be done for each target to create item-level priming effects. The resulting item-level priming effects based on the English data will be correlated (i.e., Pearson’s *rho*) with the equivalent item-level priming effects based on the German data. The point estimate of the correlation coefficient and its 95% confidence interval as well as the p-value (H0: Pearson’s *rho* = 0; H1: Pearson’s *rho* > 0) will serve as the main outcome of interest to answer the research question.

To reiterate, you will next be asked to evaluate various ways to potentially process these data. These processing choices are derived from various sources, including your input from the previous survey. On each page, you’ll see a number of thematically-clustered, data-processing choices, and we ask you to judge whether they are appropriate, or inappropriate to include, given our research question. If your choice depends on the data, we provide a representative sample here [insert link]. If you are unfamiliar with a particular option, you can select “don’t know”, but you have to provide an answer for every option: appropriate, inappropriate, or don’t know. Importantly, you can consider none, one, more than one or all options on a given page as appropriate or inappropriate. You don’t have to give a justification, though at the end of the survey you can leave comments. If you selected more than one option as appropriate, you will next be prompted to rank order the selected options from best/most preferred to worst/least preferred yet still appropriate. Note that ties are not allowed.

For example, imagine that there are four options to deal with outliers, Option A, Option B, Option C, and Option D. Imagine, you deem A, C and D appropriate and B inappropriate. After selecting your respective answers, you will be prompted to rank order Option A, C, and D by entering a number from, in this case, 1 to 3, with 1 being your best/most preferred option and 3 being your worst/least preferred option. If you deem only one or none of the options to be appropriate, you won’t be prompted to rank order them.

Note that some options may appear contradictory. For example, Option A: remove data points for reason X; Option B: keep data points regardless. You may deem only one of the two options appropriate, but that doesn’t have to be the case. You can also consider both to be appropriate in that there are sound arguments for either option, and they would both allow you to answer the research question. Alternatively, you can also consider both options to be inappropriate, if you think neither option is justifiable given the research question.

In addition, some options are subsets of others. For example, Option A: remove Response Times (henceforth RTs) below X ms; Option B: remove RTs below Y ms, with Y > X. The idea is to first evaluate whether each option in itself would be appropriate. One could, for example, consider Option B to be too strict in that it gets rid of too many observations and/or Option A to be not strict enough (it includes too many observations). Again, you could also consider both options to be appropriate, and then in the next step select which of the two you’d consider to be the best/most preferred one.