Method

Erin M. Buchanan

Last Update 2024-04-15

## 0.1 Languages

43 languages were originally identified for possible data collection based on the information available from the OpenSubtitles and subs2vec project. The following languages/dialects were translated and collected at least one participant (starred languages indicate our pre-registered minimum data collection plan): Arabic, Czech\*, Danish, German\*, Greek, English\*, Spanish\*, Farsi, French, Hebrew, Hindi, Hungarian, Italian, Japanese\*, Korean\*, Dutch, Norwegian, Polish, Portuguese\*, Brazilian Portuguese, Romanian, Russian\*, Slovak, Slovenian, Serbian, Thai, Turkish\*, Urdu, Simplified Chinese\*, and Traditional Chinese. Table X portrays a summary of each language’s data collection for number of included participants (based on the pre-registered data inclusion rules), the number of excluded participants, proportion correct answers for included participants, and median time completion for included participants. The complete breakdown of gender, education, age, and stimuli completion can be found in the supplemental materials. The following languages met the minimum data collection requirements and will be analyzed in this manuscript: Czech, Danish, German, Greek, English, Spanish, French, Hungarian, Italian, Japanese, Korean, Polish, Portuguese, Brazilian Portuguese, Romanian, Russian, Serbian, Turkish, and Simplified Chinese. The stimuli list for Portuguese and Brazilian Portuguese overlapped by 90%, and data were combined such that each unique target (unrelated and related trials) obtained the minimum number of participant answers. All data is available online, even those languages that did not meet the pre-registered minimum data collection for analysis. For each language, we also provide data checks and a summary of the number of participants, trials, items, and priming trials during data processing (see Supplemental Materials).

| Language | N Include | N Exclude | Proportion Correct | Median Time |
| --- | --- | --- | --- | --- |
| Arabic | 133 | 102 | 0.92 | 18.67 |
| Czech | 1,074 | 362 | 0.94 | 19.76 |
| Danish | 829 | 167 | 0.93 | 18.70 |
| German | 2,628 | 469 | 0.94 | 19.02 |
| Greek | 689 | 130 | 0.94 | 18.48 |
| English | 5,122 | 1,607 | 0.92 | 17.64 |
| Spanish | 1,468 | 284 | 0.94 | 18.04 |
| Farsi | 192 | 110 | 0.95 | 17.71 |
| French | 869 | 142 | 0.95 | 17.68 |
| Hebrew | 247 | 74 | 0.92 | 16.63 |
| Hindi | 1 | 2 | 0.82 | 27.39 |
| Hungarian | 718 | 180 | 0.94 | 17.94 |
| Italian | 1,085 | 142 | 0.95 | 18.10 |
| Japanese | 1,165 | 680 | 0.94 | 18.69 |
| Korean | 975 | 601 | 0.91 | 17.59 |
| Dutch | 184 | 25 | 0.93 | 17.60 |
| Norwegian | 85 | 17 | 0.93 | 20.08 |
| Polish | 1,188 | 318 | 0.94 | 19.15 |
| Portuguese | 1,178 | 332 | 0.93 | 18.25 |
| Romanian | 741 | 174 | 0.94 | 19.65 |
| Russian | 1,806 | 956 | 0.94 | 19.68 |
| Slovak | 381 | 391 | 0.94 | 18.68 |
| Slovenian | 31 | 10 | 0.95 | 18.89 |
| Serbian | 681 | 109 | 0.94 | 21.01 |
| Thai | 65 | 20 | 0.95 | 18.34 |
| Turkish | 2,218 | 790 | 0.93 | 17.83 |
| Urdu | 315 | 381 | 0.88 | 22.15 |
| Simplified Chinese | 729 | 291 | 0.93 | 17.75 |
| Traditional Chinese | 174 | 67 | 0.92 | 18.05 |

## 0.2 Ethics and Research Labs

We did not collect any identifiable private or personal data as part of the experiment. This project was approved by Harrisburg University of Science and Technology conforming to all relevant ethical guidelines and the Declaration of Helsinki, with special care to conform to the General Data Protection Regulation (GDPR; eugdpr.org). No global exclusion criteria for participating in the study was used, except for a minimum age requirement of 18 years (i.e., adult participants). Please see the analysis section below for other participant and trial level exclusion criteria related to analyses.

133 labs completed ethics documentation for data collection, and 126 labs in 41 geopolitical regions collected data (i.e., *N* >= 1) for the study. Each data collection lab obtained local ethical review (81), relied on the ethical review provided by Harrisburg University (31), or provided evidence of no required ethical review (14). The supplemental materials provide links for the IRB approvals on the Open Science Framework (OSF) and a table of participating labs with their data collection information include languages sampled, geopolitical region of the team, compensation procedure and amount, online versus in person testing, and testing type (individual participants or classroom type settings). This information can be matched to study data using the lab code that is present in the participant and trial level files.

## 0.3 Participants

35904 participants opened the link to the study, and 31624 proceeded through to complete at least one trial of the study trials (i.e., past the practice trials). Of these participants, 26971 were retained for analysis because they met our three participant level inclusion rules: 1) at least 18 years of age, 2) completed at least 100 trials, and 3) scored at least 80% correct. All exclusion criteria are summarized in the results section for clarity. The pre-registered plan calculated accuracy as in the planned scripts; however, an administrative team discussion revealed that the pre-registered report could be interpreted as . If accuracy is recalculated using this formula, 28162 participants would be included for analysis. This report will use the stricter criterion of accuracy for analysis, while the analysis using the rescored accuracy can be found in our supplemental materials. The analyses reported below only examine languages that met the minimum data criteria, which includes 32897 total participants, 29155 who completed at least one trial, 25163 who meet the strict inclusion rules, and 26197 that meet the rescored version of the inclusion rule for accuracy. Descriptive statistics about participants are provided below in the results for the 25163 participants who meet the strict inclusion rules.