|  |  |
| --- | --- |
| **Name of the file** | **Description of the file** |
| **test1\_correctResponses.csv** | This file contains the number of correct responses for test 1 split by bilingual and monolingual groups and by linguistic and non-linguistic stimuli.  Variables (columns in CSV file):  **bil\_ling\_correct:** number of correct answers given by each Basque-Spanish native speaker on linguistic stimuli  **bil\_nonling\_correct:** number of correct answers given by each Basque-Spanish native speaker on non-linguistic stimuli  **mon\_ling\_correct:** number of correct answers given by each Spanish native speaker on linguistic stimuli  **mon\_nonling\_correct:** number of correct answers given by each Spanish native speaker on non-linguistic stimuli  This file was used to run Bayesian one-sample Wilcoxon signed-rank t-test on each variable. All Bayesian tests were run in JASP v. 0.19.1 and can be reproduced using this CSV file |
| **FDRateAnalysis.csv** | This file contains the test 1 data for precision, recall and specificity for each participant.  Variables (columns in CSV file):  **part\_num:** each participant in this file is assigned an individual number  **group:** bil (sample of Basque-Spanish speakers) and mon (sample of Spanish speakers)  **hits\_TP\_ling:** number of *hits* on linguistic material (per participant). Each *hit* is a case when a word (stem or affixed) was presented during a dual-forced-choice test, and a participant responded that it is indeed a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *true positives*, hence the abbreviation **TP** in the variable name  **miss\_FN\_ling:** number of *misses* on linguistic material (per participant). Each *miss* is a case when a word (stem or affixed) was presented during a dual-forced-choice test, and a participant responded that it is not a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *false negatives*, hence the abbreviation **FN** in the variable name  **FA\_FP\_ling:** number of *false alarms* on linguistic material (per participant). Each *false alarm* is a case when a foil was presented during a dual-forced-choice test, and a participant responded that it is indeed a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *false positives*, hence the abbreviation **FP** in the variable name  **CR\_TN\_ling:** number of *correct rejections* on linguistic material (per participant). Each *correct rejection* is a case when a foil was presented during a dual-forced-choice test, and a participant responded that it is not a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *true negatives*, hence the abbreviation **TN** in the variable name  **hits\_TP\_NonLing:** number of *hits* on non-linguistic material (per participant). Each *hit* is a case when a word (stem or affixed) was presented during a dual-forced-choice test, and a participant responded that it is indeed a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *true positives*, hence the abbreviation **TP** in the variable name  **miss\_FN\_NonLing:** number of *misses* on non-linguistic material (per participant). Each *miss* is a case when a word (stem or affixed) was presented during a dual-forced-choice test, and a participant responded that it is not a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *false negatives*, hence the abbreviation **FN** in the variable name  **FA\_FP\_NonLing:** number of *false alarms* on linguistic non-material (per participant). Each *false alarm* is a case when a foil was presented during a dual-forced-choice test, and a participant responded that it is indeed a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *false positives*, hence the abbreviation **FP** in the variable name  **CR\_TN\_NonLing:** number of *correct rejections* on non-linguistic material (per participant). Each *correct rejection* is a case when a foil was presented during a dual-forced-choice test, and a participant responded that it is not a word from the language he had listened to. In terms of FDR approach, such responses are referred to as *true negatives*, hence the abbreviation **TN** in the variable name  **Precision\_ling:** precision scores on linguistic material  **Recall\_Ling:** accuracy in recognition of tokens that are true words, on linguistic material  **Specificity\_ling:** accuracy of rejecting tokens that are foils, on linguistic material  **Precision\_NonLing:** precision scores on non-linguistic material  **Recall\_NonLing:** accuracy in recognition of tokens that are true words, on non-linguistic material  **Specificity\_NonLing:** accuracy of rejecting tokens that are foils, on non-linguistic material  \*-\*-\*-\*-\*  **False** and **true negatives** and **false** and **true positives** are used to calculate **precision** (i.e., general accuracy) in recognition test (i.e., the proportion of words among endorsed tokens) **Precision** measure reflects how well people endorse the words and reject the foils, achieving higher number of correct responses.  Precision is a product of **recall**, which is also referred to as *sensitivity* (proportion of endorsed words among all presented words, which reflects how well people can endorse the true words), and **specificity** (proportion of foils that were not endorsed, which reflects how well people can reject the false items).  As the same accuracy level can be achieved by a higher level of recognition of true items in one group and high level of rejecting fake items (foils) in another group of participants, we used **recall** and **specificity** separately, in addition to the **precision** measure. |
| **Test\_2\_PrefixSufixPreference.csv** | This file contains data for test 2, in which participants listened to a pair of tokens, one token was a prefixed sequence and the other was a suffixed sequence, both sequences were legitimate and were embedded into the familiarization stream multiple (and an equal number of) times. Participants had to choose the one token from the pair – either prefixed or suffixed – which, in their opinion, was more likely to be a word from the language they had listened to. This test was administered to find out whether there is a cognitive bias towards (preference for) selecting one type of the sequences over the other, and whether the cognitive bias is affected by typological bias present in the native language(s) of participants.  **part\_num:** each participant in this file is assigned an individual number. The same number is assigned to the participant across test 1 (in **FDRateAnalysis.csv** file) and in test 2 (**Test\_2\_PrefixSufixPreference.csv** file)  **group:** bil (sample of Basque-Spanish speakers) and mon (sample of Spanish speakers)  **pr\_YES\_Ling:** number of times a participant chose a prefixed sequence on linguistic material  **suff\_YES\_Ling:** number of times a participant chose a prefixed sequence on linguistic material  *NB:* the sum of pr\_YES\_Ling and suff\_YES\_Ling is six (this is number of test trials)  **pr\_YES\_NLing:** number of times a participant chose a prefixed sequence on non-linguistic material  **suff\_YES\_NLing:** number of times a participant chose a prefixed sequence on non-linguistic material  *NB:* the sum of pr\_YES\_NLing and suff\_YES\_NLing is six (this is number of test trials) |