

Methods

Updated: 01/July/2021

Participants

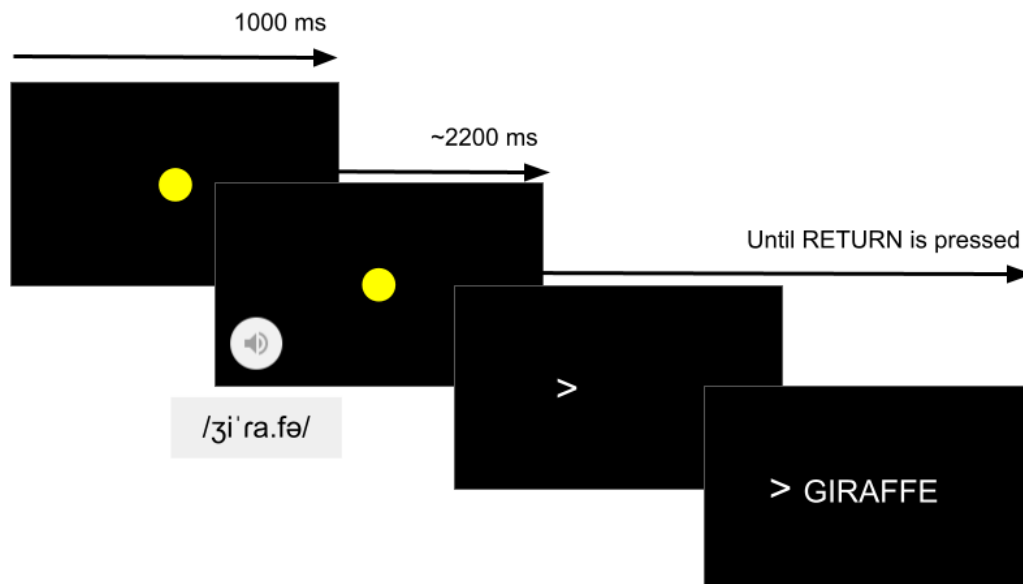
Data collection took place from 04th June, 2020 to 28th June, 2020. We collected data from 105 participants ($M_{age} = 21.79$, $SD_{age} = 21.79$, $Range_{age} = 21.79-33$). 73 participants were British English native speakers living in United Kingdom (47 female), and 32 participants were Spanish native speakers living in Spain (27 female). Participants in UK were recruited via Prolific (5£ compensation) and SONA (compensation in academic credits). Participants in Spain were contacted via announcements in Faculties, and were compensated 5€ or an Amazon voucher for the same value. Participants were asked to complete the experiment in a quiet place with good internet connection. We excluded data from participants that a) self-rated their oral and/or written skills in a second or third language as higher than 4 in a 5-point scale ($n = 1$), b) were diagnosed with a language ($n = 2$)¹, or c) did not contribute more than 80% of valid trials ($n = 9$).

Procedure

The experiment was implemented online using Psychopy/Pavlovia ([peirce2019?](#)). Participants accessed the study from a link provided by Prolific or SONA and completed the experiment from a browser (Chrome or Mozilla). First, participants were informed about the aims of the study and gave informed consent for participating. Second, participants answered a series of questions about their demographic status, their language background, and the set up they were using for completing the study. Third, participants completed the experimental task. Participants were informed that they would listen to a series of pre-recorded words in Catalan or Spanish (English participants) or only Catalan (Spanish participants). They were instructed to listen to each word, guess its meaning in English (English participants) or Spanish (Spanish participants), and type their answer as soon as possible. English participants were randomly assigned to the list of Catalan or Spanish trials. Participants in the Catalan list were presented with 86 trials, and participants in the Spanish list were presented with 103 trials.

Each trial started with a yellow fixation dot presented during one second on the centre of the screen over a black background. After one second, the audio started playing while the dot remained being displayed until the audio ended. Upon the offset of the fixation point and audio, participants were prompted to write their answer by a “>” symbol. Typed letters were displayed in the screen in real time to provide visual feed-back to participants. Participants were allowed to correct their answer. Then, participants pressed the RETURN key to start and new trial. We excluded trials where participants did not type an existing word in the correspondent language, or did not type anything at all. Trials where the response was mistyped by only one character were counted as correct, as long as the respond did not correspond to a distinct word. Participants contributed a total of 8077 valid trials (5235 in Catalan, 2842 in Spanish). The task took approximately 15 minutes to be completed.

¹We originally planned to exclude participants that reported any visual impairment that glasses would not correct. This item was phrased as “Do you have normal or corrected-to-normal VISION? (Yes/No)” in English, and as *¿Tienes problemas de VISIÓN que unas gafas o lentes de contacto NO corrijan? (Sí/No)*. However, the proportion of Spanish participants that reported visual impairment was implausibly large ($n = 6$, 18.75%). This is possibly due to these participants using glasses daily and not having read the item until the end, where it is indicated that the use of glasses is considered as normal vision.



Stimuli

Participants listened to one audio file in each trial. This audio file corresponded to a word in Catalan (for Spanish speakers, and for English participants allocated in the Catalan condition) or Spanish (for English speakers allocated in the Spanish condition). The audio files were the same ones used in child experiments conducted in the Laboratori de Recerca en Infància of Universitat Pompeu Fabra (Barcelona, Spain). These audio files were recorded by a proficient Catalan-Spanish female bilingual from the Metropolitan Area of Barcelona in a child-directed manner. Catalan and Spanish words were recorded at 44,100 Hz in separate files in the same session, and then de-noised using Audacity and normalised at peak intensity using Praat (boersma2018?).

Lexical frequencies in Catalan and Spanish were extracted from SUBTLEX-CAT (boada2013?) and SUBTLEX-ESP (cuetos2011?), respectively, and are expressed as Zipf scores (van-heuven2012?; zipf1949?). The average lexical frequency (counts per million) of the words was 4.29 ($SD = 0.59$, $Min = 2.87$, $Max = 6.14$). The average lexical frequency of the Catalan words was 4.31 ($SD = 0.59$, $Min = 2.87$, $Max = 5.68$), and the average lexical frequency of the Spanish words was 4.32 ($SD = 0.56$, $Min = 2.87$, $Max = 5.68$).

Data analysis

We divided participants in three groups: English natives tested in Catalan (ENG-CAT), English natives tested in Spanish (ENG-SPA), and Spanish natives tested in Catalan (SPA-CAT). We analysed the accuracy of responses to each item by calculating the proportion of correct responses across participants, resulting in one data point per item and group. We compared how our similarity score fitted the data in comparison with other measures: total shared phonemes, shared phonemic onset, and orthographical Damerau-Levenshtein distance. We fit several Bayesian linear models (sorensen2015?), each including one of the aforementioned variables as predictor, and the proportion of correct responses as output. Continuous predictors were standardised and categorical predictors were sum-coded (schad2020?). Missing predictor scores were imputed before fitting the models using multiple imputation (vanbuuren2013?). To test and account for cross-group differences, we included a random intercept for each group. We compared models using leave-one-out

cross-validation (LOO, (vehtari2016?)). More information about the models and model comparison can be found in Appendix X.

All analyses were performed in R environment (rcoreteam2019?). We used the `tidyverse` family of R packages to process data and to generate figures, the `mice` R package (vanbuuren2013?) for multiple imputation of missing data, and the `brms` R package (burkner2017?) to fit and compare models, which uses the Stan probabilistic programming language (carpenter2016?).