22-09-2025
Suzan Dilara Scheffer

MODELLING OF HUMAN LANGUAGE
ASSIGNMENT 2

Bilingual Language Processing - Categorical Data Analysis

In this assignment, you will be analyzing categorical data collected from a group of bilinguals, namely Heritage Language Speakers. We talked about this group of speakers during our lecture.

Learning Goals for this lab:

- Understand how an experiment is set up to study language processing
- Learn how to analyze experimental results from human subjects, particularly with a categorical dependent variable
- Write up research results in article form

Practical information:

- We are using R to analyze our data. We will use RStudio as the user interface for R.
- The files needed are on GitHub. The link is on BS.
- Write the lab report as if you are writing a conference article.
- The reports shouldn't be longer than 2-3 pages.
- The deadline is October 13th. If you require additional time, please let me in advance.
- This assignment accounts for 30% of your final grade.

Looking at the data:

You need two data sets in a .csv file called:

Assg_2_participant_data.csv

Assg_2_experiment_output.csv

The experiment output is from a self-paced reading study we conducted in order to examine the linguistic phenomenon known as evidentiality, which refers to the indication of the source or basis of information within an utterance. Evidentiality is the marking of the source of information, thereby indicating how one gained the knowledge regarding the event that is referred to in one's statement (Aikhenvald, 2004; Chafe & Nichols, 1986; De Haan, 1999; Faller, 2006; Lazard, 2001; Willett, 1988).

Our primary objective was to investigate how heritage language speakers process sentences with evidentiality markers in a self-paced reading experiment, while also examining the influence of various language background factors, such as age of acquisition and proficiency, on their judgments regarding sentence acceptability.

We manipulated two factors: grammaticality and evidentiality. Grammaticality has two

levels, with "mismatch" representing ungrammatical sentences and "match" representing grammatical sentences. Evidentiality, on the other hand, involved two different evidentiality markers, "direct" and "indirect," resulting in four distinct conditions:
SD (Seen-Direct: Matching Direct), HI (Heard-Indirect: Matching Indirect), HD* (Heard-Direct: Mismatch Direct), SI* (Seen Indirect: Mismatch Indirect). See the *stimuli example* for example sentences. For a comprehensive list of experimental stimuli, please refer to the repository: https://github.com/sdtokac/Chapter-4_selfpacedreadingtask_stimuli

The participants had to evaluate the sentences they read with one of two possible interpretations by choosing a smiling emoji (🙂) if they thought the sentence was good and free from any errors OR a sad one (😕) if they thought something was off with the sentence. The answers were recorded and annotated as 1 for correct answers and 0 for wrong answers.

Working through the R file
The first thing you should do is work through the R file analysis. We've given tasks and some explanations in the script, and you need to write down formulas, run them, and report the outputs. Make sure you understand what is going on, and in the end, you should get the best, maximal model. You should also understand what the random factors and fixed factors[1] in this experiment are and what differences were significant.

Note: The script is very detailed because the aim of this assignment is not to teach you how to do statistical analyses but how to conduct research with human subjects, which variables are relevant, and how to report them correctly (to show understanding).

Doing an independent analysis
What you have to do: Do the analysis. Find out the descriptive stats. Use glmer to determine the best, maximal model that you can get to converge, also using random slopes. Generate appropriate visuals. Make sure you understand what happened in the data and what your discovery was. Write a lab report with results. But remember everything needs a story so you have to write an Introduction with reasoning behind this study. Check the grading guidelines to make sure you included all the relevant information on your report.

[1] Fixed effects and Random effects: Fixed effects are not affected by increasing the sample of participants. So if one of the factors is the Grammaticality, even if you test more participants, this factor will still only have two levels: match and mismatch. Random effects are generally Participant and Item in linguistic experiments. These factors are random because you can't predict beforehand what levels they will contain and what variation they will show.