Data Visualisation and Statistics for Language Sciences

*HT2021*

**Final assignment: Free-listing/Exemplar-listing**

You will conduct two different experiments: one free-listing task in which participants will be asked to list as many COLOURS, TASTES, and SMELLS as they can; and one exemplar-listing task in which participants will be asked to list as many examples of specific TASTES and SMELLS as they can.

**Analyses**

***Experiment 1***

The analysis of your free-listing data will serve two purposes: you will need to find out which TASTE and SMELL terms to include in the second experiment, but you will also need to do some more analyses for the assignment itself.

First, you need to choose stimuli for Experiment 2. To do this, create a frequency table of all unique terms listed in each individual domain—make sure you correct any writing errors. Discard all terms that were only listed once. Then, for each term that participants listed, code it as one of the following categories:

1. ***Domain-specific*** (abstract terms that refer to a perceptual quality), e.g., *red*, *green*, *blue* for COLOUR; *sweet*, *sour*, *bitter* for TASTE; and *aromatic*, *stinky*, *musty* for SMELL
2. ***Source-based*** (terms that refer to a specific source), e.g., *gold*, *silver* for COLOUR; *sugary*, *creamy* for TASTE; and *flowery*, *fruity* for SMELL
3. ***Evaluative*** (terms that give some evaluative judgement), e.g., *pretty* for COLOUR; *delicious* for TASTE; and *pleasant* for SMELL
4. ***Other*** (terms that don’t fit any of the other groups, or terms that are from other domains), e.g., *bright*, *dark* for COLOUR; *mild*, *dry* for TASTE; and *fresh*, *sour* for SMELL

For TASTES and SMELLS, select the most frequent domain-specific terms for use in Experiment 2. There will usually be a clear cut-off point between frequent and infrequent terms. Many languages will not have highly frequent domain-specific terms for SMELLS, so you might want to choose different types of terms, or omit them all together. You can sit together with the instructor to make a final selection of terms for the follow-up exemplar-listing task.

Finally, for the assignment, you will also write a short report on the results of the free-listing task, which should focus a comparing the three domains (COLOUR, TASTE, SMELL) on:

-How many terms people listed on average per domain

-How many unique terms people listed in total for each domain

-The ratio of domain-specific/source-based/evaluative/other terms people listed in the three domains, for both *types* (unique responses) and *tokens* (all responses)

Feel free to add any additional analyses of your own.

Use informative visualisations to present your results!

***Experiment 2***

The analysis of your exemplar-listing data will only be for the assignment itself as there will be no more follow-up experiments after this. As with the free-listing, you will write a short report summarising the findings.

Create a frequency table of all examples that people listed and discard all examples that were only listed once. Make sure you correct any writing errors. You might also want to collapse some responses into a single category, e.g., the responses ‘jalapeño’, ‘jalapeños’, and ‘jalapeño peppers’ can all be coded into a single response ‘jalapeño’.

Compare the different TASTE and SMELL terms based on how many examples people listed on average per terms, how many unique examples people listed in total for each term, and compare the type-token ratio for each term.

Finally, you will be given some comparable data for Dutch TASTE terms. Using some type of distance measure to calculate how similar people experience taste terms referring to bitter *bitter*, *salty*, *sour*, and *sweet* in Dutch and in the language in which you collected your data. There are several options for calculating such distances, e.g., *Jaccard Index* based on the unique examples, *cosine similarity* based on all the examples listed, the *listing order* method (e.g., as used in Medin et al. 2010), or a different measure of your choice. Pick one, motivate your choice and report the results.

Feel free to add any additional analyses of your own.

Use information visualisations to convey your findings, especially for the distance measure!