Grant Number: ES/R00837X/1

Sponsor: ESRC

Project title: Optimal Categorisation: the origin and nature of gender from a psycholinguistic perspective

Project Website: https://nominal-categorisation.surrey.ac.uk/

This document describes the data from the Optimal Categorisation project that has been archived with the UK Data Service. The data is described according to which folder it belongs in. The data is sorted according to the six different experiments and participant metadata.

1. Card Sorting Experiment
2. Eye-tracking Experiment
3. Free-listing Experiment
4. Possessive Labelling Experiment
5. Storyboard Experiment
6. Video Vignettes Experiment
7. Participant Metadata

1. Card Sorting Experiment

To investigate the semantic domains of the possessive classifiers in the languages a card sorting experiment was used, which enables us to get an insight into how participants categorise objects and what governs perceived similarity. Participants sorted images in a free card sort task, where they were asked to group the images in any way they chose, followed by a structured card sort task, where they were asked to group images based on the classifiers in their language. All participants were presented with the same stimuli, to enable comparison of sorting and the relative influence of classifiers across the different languages. Comparison between the groupings in the free sort task and the structured sort task provides an insight into the relative influence of linguistic and cultural factors on participants' categorisation. Data were analysed using a mixed methods approach.

Participants were presented with 60 cards detailing a standardised set of images. These images depicted entities and different interactional uses of an entity. For example, a live pig and a pig being roasted. Participants first completed a naming task. They were presented with each image in turn and asked to name the entity to ensure understanding. The images were then laid out in front of the participants, who were asked to free sort the cards into groups. After the task, participants were asked to give a summary of each pile, explaining why these cards were grouped together. The cards were then shuffled again and laid out in front of the participants who were then asked to sort the cards into groups relating to the classifiers of their mother-tongue language, in a structured card sort. This ensured that the different classifiers were used. Participants were asked to sort images into groups according to which word they would use to mean 'mine'.

Publications:

Grandison, A., Franjieh, M., Greene, L., & Corbett, G. G. (2021). Optimal categorisation: the nature of nominal classification systems. Cadernos de Linguística, 2(1). https://doi.org/10.25189/2675-4916.2021.V2.N1.ID393

The following files have been archived for the Card Sorting Experiment in the OC-Card-sorting Zip file:

Image_descriptions.csv	CSV file. Descriptions of the images used in the card sorting experiment. Numbers in the csv file correspond to the file name in the Images.zip folder
Card-sort-protocol-.pdf	PDF file. Experimental protocol and instructions for participants
Images.zip	Zip folder containing 60 jpg images used in the card sorting experiment. File numbers 1-60 correspond the the numbers in Image_descriptions.csv.
Data>Iaai folder	Contains 16 excel files relating to the 16 participants who took part in the card sorting experiment from the Iaai language in New Caledonia. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task. An additional Card Sort Readme.pdf explains any missing data.
Data>Lewo folder	Contains 23 excel files relating to the 23 participants who took part in the card sorting experiment from the Lewo language in Vanuatu. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task.
Data>Merei folder	Contains 21 excel files relating to the 21 participants who took part in the card sorting experiment from the Merei language in Vanuatu. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task. An additional Card Sort Readme.pdf explains any missing data.
Data>Nelemwa folder	Contains 12 excel files relating to the 12 participants who took part in the card sorting experiment from the Nêlêmwa language in New Caledonia. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the

	results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task. An additional Card Sort Readme.pdf explains any missing data.
Data>Rral folder	Contains 23 excel files relating to the 23 participants who took part in the card sorting experiment from the Rral/North Ambrym language in Vanuatu. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task.
Data>Vatlongos folder	Contains 24 excel files relating to the 24 participants who took part in the card sorting experiment from the Vatlongos language in Vanuatu. Each excel file contains the results from each of the three parts of the experiments in sheets 1-3. Sheet 1 has the results from the image naming task. Sheet 2 has the results from the free sort task. Sheet 3 has the results from the classifier sort task.

2. Eye-Tracking Experiment

In order to determine how speakers process auditory data, eye tracking was employed. By evaluating fixations and scan patterns in relation to particular regions of interest, eye tracking offers objective measures of automatic processing to identify patterns in attention. This methodology will give precise insights into the categorisation-related cognitive processes. Participants were first presented with a center attention grabber before a pair of images that were congruent or incongruent with that classifier in terms of usage and number was displayed on the left and right of the screen. Eye tracking was used to answer the question: Are decision times longer for classifiers with more complicated and diverse semantic domains?

The experiments had three tasks:

Part 1: All participants were first familiarised with the visual stimuli. The images were presented twice, with simultaneous audio recordings of the name of the image in the participant's language.

Part 2: All participants completed an eye calibration before the experiment began. The experiment started with an auditory cue (a possessive classifier) during the whirligig attention grabber portion before an image was displayed. An audio classifier was played when the image appeared, either congruent or incongruent to one of the visual objects. A message on the screen encouraged participants to click the mouse after three seconds. The next audio classifier cue was presented for two seconds, and the next image of two pairs of images appeared. Up until the end of all the trials, this was repeated.

Part 3: Finally, participants were presented with each image and a possessive classifier. Participants had to state whether the classifier they heard is congruent with the image.

The experiment was conducted with participants from five languages – Iaai, North Ambrym (Rral), Merei, Vatlongos and Lewo. The eye-tracker malfunctioned during the trials with Nêlêmwa speakers and only partial data is included for four participants.

The following files have been archived for the Eye-Tracking Experiment in the OC-Eye-tracking zip folder:

Eye-tracking-protocol.pdf	PDF file. Experimental protocol and instructions for participants in English, Bislama and French
Material>Part 1 folder	• 8 jpg images used in part 1 of the eye-tracking experiment. • 40 wav audio files, saying the name of each image, spoken in the languages Iaai, Lewo, Merei, Rral, and Vatlongos. • 10 e-prime files for (two for each language) for part 1 of the experiment. To be opened with the e-prime 3 program
Material>Part 2 folder	• 5 excel spreadsheets relating to the full data export from Tobii-Pro-Lab for each of the languages Iaai, Lewo, Merei, Rral, and Vatlongos for part 2 of the experiment. • the folder Congruence-2022-project export folder which contains the Tobii Pro Lab project which can be re-imported.
Material>Part 3 folder	• 8 images used in part 3 of the experiment • 16 audio files saying the name of each classifier in the languages Iaai, Lewo, Merei, Rral, and Vatlongos. • 10 e-prime files for (two for each language) for part 3 of the experiment. To be opened with the e-prime 3 program • 5 folders containing txt and csv file outputs from e-prime 3 of each of the participant's answers from part 3 in each language. A notes txt file shows any changes from the original data (found in the data folder)
Data>Iaai folder	Contains 19 zip folders relating to the 19 participants who took part in the eye-tracking experiment from the Iaai language in New Caledonia. The data mainly concerns part 3 of the experiment. Each folder contains the following data collected from each participant. P01-P19 contains:

	• A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment Additionaly, P01 contains: • A .tsv file for the calibration results from part-2 of the eye-tracking experiment from tobii pro lab • A .tsv file of full data export of part-2 of the experiment from tobii pro lab • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment
Data>Lewo folder	Contains 22 zip folders relating to the 22 participants who took part in the eye-tracking experiment from the Lewo language in Vanuatu. The data mainly concerns part 3 of the experiment. Each folder contains the following data collected from each participant: P01-P22 contains: • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment P10 also contains (erroneously) data pertaining to the possessive labelling experiment: • 138 .wav audio file responses to the possessive labelling stimulus • A .edat3 e-prime 3 data file with the participant responses for the possessive labelling experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for the possessive labelling experiment
Data>Merei folder	Contains 21 zip folders relating to the 21 participants who took part in the eye-tracking experiment from the Merei language in Vanuatu. Each folder contains the data

	collected from each participant for part-3 of the experiment. P01-P21 contains: • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment
Data>Nelemwa folder	Contains 4 zip folders relating to the 4 participants who successfully took part in the eye-tracking experiment from the Nêlêmwa language in New Caledonia. . Each folder contains the data collected from each participant for part-3 of the experiment. An additional Read me file.pdf explains the lack of data from this sample. P01-P04 contains: • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment
Data>Rral folder	Contains 22 zip folders relating to the 22 participants who took part in the eye-tracking experiment from the Rral/North Ambrym language in Vanuatu. . Each folder contains the data collected from each participant for part-3 of the experiment. P01-P22 contains: • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment
Data>Vatlongos folder	Contains 22 zip folders relating to the 22 participants who took part in the eye-tracking experiment from the Vatlongos language in Vanuatu. Each folder contains the data

	collected from each participant for part-3 of the experiment. P01-P22 contains: • A .edat3 e-prime 3 data file with the participant responses for part 3 of the experiment • A .txt file converted from the original .edat3 file for opening in excel etc. • An xml file advisory report generated by e-prime 3 for part 3 of the experiment

3. Free listing experiment

Free-listing is a technique used to isolate and define a domain and explore salient members of a domain. Quite simply, participants are asked to list members of a domain or category specified by the researcher. Lists collated from multiple participants can be combined to uncover salient members of a category. The frequency of an item across lists produced by different participants and the order of an item within individual participants' lists have been shown to be indicators of the salience of a category member. These metrics are often combined to give an overall numeric value indicating salience, such as Smith's Salience or the Cognitive Salience Index.

Participants from the six languages heard a recording of each of the possessive classifiers in their language. After each classifier, participants had three minutes to say as many nouns that were associated with that classifier.

Publications:

Franjieh, M., Grandison, A., Dotte, A-L., & Corbett, G. G. 2022. Implementing free-listing: possessive classifiers in Oceanic. Semantic Fieldwork Methods, 4(2).

The following files have been archived for the Free Listing Experiment in the OC-Free-Listing Zip folder:

Free-list-protocol.pdf	PDF file. Experimental protocol and instructions for participants in English, Bislama and French
Freelist-audio.zip	Zip folder containing audio files of each classifier in each language included in the study for participants to listen to. • Iaai sub-folder contains .wav recordings for 23 classifiers. • Nêlêmwa sub-folder contains .wav recordings for 20 classifiers • Rral sub-folder contains .wav recordings for 5 classifiers (for Rral/North Ambrym language) • Vatlongos sub-folder contains .wav recordings for 4 classifiers

	• Lewo sub-folder contains .wav recordings for 3 classifiers • Merei sub-folder contains .wav recordings for 2 classifiers
Data>laai folder	Contains 17 Excel files relating to the 17 participants who took part in the free-listing experiment from the laai language in New Caledonia. Each file contains the data collected from each participant.
Data>Lewo folder	Contains 23 Excel files relating to the 23 participants who took part in the free-listing experiment from the Lewo language in Vanuatu. Each file contains the data collected from each participant.
Data>Merei folder	Contains 22 Excel files relating to the 22 participants who took part in the free-listing experiment from the Merei language in Vanuatu. Each file contains the data collected from each participant. Also contains a pdf file flagging a participant whose data has been excluded from the experiment.
Data>Nelemwa folder	Contains 11 Excel files relating to 11 of the 13 participants who took part in the free-listing experiment from the Nelemwa language in New Caledonia. Each file contains data collected from each participant. Also contains a pdf file flagging 2 participants (participant 1 and participants 12) whose data was excluded from the experiment.
Data>Rral folder	Contains 23 Excel files relating to the 23 participants who took part in the free-listing experiment from the Rral language in Vanuatu. Each file contains data collected from each participant.
Data>Vatlongos folder	Contains 24 Excel files relating to the 24 participants who took part in the free-listing experiment from the Vatlongos language in Vanuatu. Each file contains data collected from each participant.

4. Possessive Labelling

The possessive labelling experiment had four aims (i) to test for classifier overlap, (ii) to test for speaker variation, (iii) to map semantic domains, and (iv) to test atypical possessions.

Participants completed the experiment using a laptop, in which they listened to 140 names for different things in their language. Participants heard a beep, followed by an audio cue of the target noun, and were asked to say that the referent of the noun belonged to them. The researcher

emphasised that there are various ways to express things belong to them, but they should respond with the first way they thought. Participants were given the example: If you hear the word 'flying fox' you must say 'mine'. Or if you hear the word 'trousers' you must say 'mine'. There may be different ways to say 'mine' for these things in your language. A short three-word practice task was conducted with all participants. The researcher wrote down their answer and any comments.

The experiment used e-prime 3 and a chronos box which records audio responses and reaction times linked to the audio responses.

The following files have been archived for the Possessive Labelling Experiment in the OC-Possessive-labelling Zip folder:

Possessive-labelling-protocol.pdf	PDF file. Experimental protocol and instructions for participants in English, Bislama and French
Data>Iaai folder	Contains 20 folders. • The folder 'Iaai-Experiment' contains 140 audio clips of stimulus nouns used in the experiment, and 2 ebs files for opening the experiment in e-prime 3. • The remaining 19 folders relate to the 19 participants who took part in the possessive labelling experiment from the Iaai language. Each of these folders contains: o a zip file containing 140 .wav files of the audio responses for that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 o a txt file export of the data collected.
Data>Lewo folder	Contains 23 folders. • The folder 'Lewo-Experiment' contains 140 audio clips of stimulus used in the experiment and 2 ebs files for opening the experiment in e-prime 3. • The remaining 22 folders relate to the 22 participants who took part in the possessive labelling experiment from the Lewo language. Each of these folders contains: o a zip file containing 140 .wav files of the audio responses for that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 o a txt file export of the data collected.

Data>Merei folder	Contains 22 folders. • The folder 'Merei-Experiment' contains 140 audio clips of the stimulus used in the and 2 ebs files for opening the experiment in e-prime 3. • The remaining 21 folders relate to the 21 participants who took part in the possessive labelling experiment from the Merei language. Each of these folders contains: o a zip file containing 140 .wav files of the audio responses for that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 o a txt file export of the data collected.
Data>Nelemwa folder	Contains 20 folders. • The folder 'Nelemwa-Experiment' contains 140 audio clips of the stimulus used in the experiment and 2 ebs files for opening the experiment in e-prime 3. • The remaining 19 folders relate to the 19 participants who took part in the possessive labelling experiment from the Nelemwa language. Each of these folders contains: o a zip file containing 140 .wav files of the audio responses for that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 o a txt file export of the data collected.
Data>Rral folder	Contains 23 folders. • The folder 'Rral-Experiment' contains 141 audio clips of the stimulus used in the experiment and 2 ebs files for opening the experiment in e-prime 3. • The remaining 22 folders relate to the 22 participants who took part in the possessive labelling experiment from the Rral language. Each of these folders contains: o a zip file containing 140 .wav files of the audio responses for

	that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 ○ a txt file export of the data collected.
Data>Vatlongos folder	Contains 23 folders. • The folder 'Vatlongos-Experiment' contains 140 audio clips of the stimulus used in the experiment and 2 ebs files for opening the experiment in e-prime 3. • The remaining 22 folders relate to the 22 participants who took part in the possessive labelling experiment from the Vatlongos language. Each of these folders contains: ○ a zip file containing 140 .wav files of the audio responses for that participant, an .edat3 data file for e-prime 3, a .txt export of the e-prime 3 data and an xml run-time advisory report generated by e-prime 3 ○ a txt file export of the data collected.

5. Storyboards Experiment

The storyboard experiment had three objectives (i) to test classifier overlap, (ii) to reveal anaphoric use of the classifiers in the Oceanic languages, and (iii) to reveal the nature of classifier systems. With intentions for two primary uses: to collect narrative recordings recounting the events depicted in the storyboards, and to create basic literacy materials for the communities in the South Pacific. This will involve a limited print run of the storyboards and will be gifted to the local schools to help with vernacular literacy projects.

Participants were presented with eight short stories, each with four pictures. The researcher began by informing participants that each narrative will be explained in Bislama and then asked them to retell the story in their own language. The researcher also asked each participant to tell them how they would claim the object in the picture belonged to the man or woman in the picture. Participants were also told that after the first picture in the story containing four pictures, they did not need to repeat the name of the object after picture one. They were simply instructed to use the term that means "his" or "hers".

The following files have been archived for the Storyboard Experiment in the OC-Storyboards Zip folder:

Storyboards-protocol-.pdf	PDF file. Experimental protocol and instructions for participants in English, Bislama and French

Storyboard images.zip	A zip file containing 32 image files used in the storyboard task grouped by story.
Data>Iaai folder	Contains 19 zip folders relating to the 19 participants who took part in the storyboard task from the language Iaai. Each zip folder contains: • 8 word documents of the data produced from each story. Each wordfile has the story in French first and the Iaai translation.
Data>Lewo folder	Contains 22 zip folders relating to the 22 participants who took part in the storyboard task from the language Lewo. Each zip folder contains: • 8 word documents of the data produced from each story. Each wordfile has the story in Bislama first and the Lewo translation.
Data>Merei folder	Contains 21 zip folders relating to the 21 participants who took part in the storyboard task from the language Merei. Each zip folder contains: • 8 word documents of the data produced from each story. Each wordfile has the story in Bislama first and the Merei translation.
Data>Nelemwa folder	Contains 19 zip folders relating to the 19 participants who took part in the storyboard task from the language Nelemwa. Each zip folder contains: • 8 word documents of the data produced from each story. Each wordfile has the story in French first and the Nêlêmwa translation.
Data>Rral folder	Contains 22 zip folders relating to the 22 participants who took part in the storyboard task from the language Rral. Each zip folder contains: • 8 word documents of the data produced from each story. Each wordfile has the story in Bislama first and the Rral/North Ambrym translation.
Data>Vatlongos folder	Contains 22 zip folders relating to the 22 participants who took part in the storyboard task from the language Vatlongos. Each zip folder contains: • 8 word documents of the data produced from each story. Each

	wordfile has the story in Bislama first and the Vatlongos translation.

6. Video Vignettes

Video vignettes were used to investigate three key criteria (i) the amount of overlap between noun and classifier, (ii) the semantic (mis)match between verb and classifier, and (iii) the typicality of contextual interaction.

Participants of the 6 different languages watched the same 24 video vignettes; these were designed to investigate the use of classifiers when the items depicted in the vignettes are used in different interactional contexts. Participants were asked to describe in one simple sentence what the actor was doing with their possessions, thus evoking a possessive classifier. Since the classifier inventories vary across the sample languages, we tested their three main classifier categories – GENERAL, DRINK, and FOOD.

Publications:

Franjieh, M., G. G. Corbett & A. Grandison. 2021. Uncovering variation in classifier assignment in Oceanic. In ExLing 2021: Proceedings of 12th International Conference of Experimental Linguistics, 11-13 October 2021, Athens, Greece (pp. 81-84). ExLing Society. https://doi.org/10.36505/ExLing-2021/12/0021/000494

The following files have been archived for the Video Vignettes Experiment in the OC-Video-Vignettes Zip folder:

Video-Experiment-Protocol.pdf	PDF file. Experimental protocol and instructions for participants in English, Bislama and French.
Video Vignettes descriptions.csv	csv file. Descriptions of the 24 video vignettes used in the experiment.
Vignettes.zip	Zip folder containing 24 mp4 files. Each file is one of the 24 video vignettes used in the experiment. The descriptions are in the Video Vignettes descriptions.csv file
Data>Iaai folder	Contains: • 16 Excel files relating to the 16 participants who took part in the video vignettes experiment from the Iaai language in New Caledonia. Each file contains data collected from each participant. • Also contains pdf file 'Video Vignette Readme' flagging a participant (participant 10) who did not take part in the experiment.
Data>Lewo folder	Contains: • 23 Excel files relating to the 23 participants who took part in the video vignette experiment from the Lewo language in Vanuatu. Each file contains

	the data collected from each participant.
Data>Merei folder	Contains: • 21 Excel files relating to the 21 participants who took part in the video vignette experiment from the Merei language in Vanuatu. Each file contains the data collected from each participant. • Also contains pdf file 'Video Vignette Readme' flagging a participant (participants 12) who did not complete the experiment.
Data>Nelemwa folder	Contains: • 11 Excel files relating to 11 of the 13 participants who took part in the video vignettes experiment from the Nelemwa language in New Caledonia. Each file contains data collected from each participant. • Also contains pdf file 'Video Vignette Readme' flagging 2 participants (participant 1 and participant 12) whose data was excluded from the experiment.
Data>Rral	Contains: • 23 Excel files relating to the 23 participants who took part in the video vignettes experiment from the Rral language in Vanuatu. Each file contains data collected from each participant.
Data>Vatlongos	Contains: • 24 Excel files relating to the 24 participants who took part in the video vignettes experiment from the Vatlongos language in Vanuatu. Each file contains data collected from each participant.

7. Participants Metadata Folder

The following participant metadata files have been in the OC-Participant-Metadata Zip folder:

Eyetracking, Possessive labelling and Storyboards 2022 folder	Contains 6 csv files, one for each of the languages Iaai, Lewo, Merei, Nelemwa, Rral, Vatlongos. Each file contains the demographic

	information (gender, age, education level) for the participants who took part in the eye tracking, possessive labelling and storyboard experiments.
Pile sorting, Free listing and Video vignettes 2019 folder	Contains 6 csv files, one for each language, included in the experiments; Iaai, Lewo, Merei, Nelemwa, Rral and Vatlongos. Each file contains demographic information (gender, age, education level) for the participants who took part in the free listing and video vignette experiments.