

The emergence of grammatical gender from possessive classifiers: an experimental study

Data and Analysis Supplementary Materials

This file details the data from the storyboard experiment, the Python code to transform the data, and the R file that produces the analysis in the paper submitted to Glossa in April 2024.

There are six Oceanic languages in our sample. Four from Vanuatu (Merei, Lewo, Vatlongos, and North Ambrym). There are two languages from New Caledonia (Nêlêmwa and Iai).

Note – often in the data files the language North Ambrym is referred to as Rral, which is the endonym for the language. In the paper, the language is referred to by its exonym, North Ambrym as it is more commonly known.

The following groups of files are described below in detail:

Choosing Nouns File

Experiment Participant Files

Data preparation for individual languages

Combining individual language data

R Statistical Analysis

Choosing Nouns File

The Storyboard_nouns_choice.csv file contains information concerning the matching nouns across the North Ambrym (Rral) and the Vatlongos corpus; how frequent the nouns were in the initial free-list experiment, along with the Cognitive Salience Index (CSI) score; normalised frequency for the two corpora (per 100,000); frequency quartile and average CSI rank. The 'DELETE' column indicates that the items were not used in the final set for choosing the eight nouns – see paper for details.

Experiment Participant Files

Each language has a number of data files which were created by copying the answers participants gave onto a spreadsheet. These were then coded for which classifier/possessive construction was used. The number of data files for each language depends on the number of participants for each language

Language	Number of participant files	Filename examples
Merei	21	Merei-Storyboards-P01.csv, Merei-Storyboards-P21.csv
Lewo	22	Lewo-Storyboards-P01.csv, Lewo-Storyboards-P22.csv
Vatlongos	22	Vatlongos-Storyboards-P01.csv, Vatlongos-Storyboards-P22.csv
North Ambrym / Rral	22	Rral-Storyboards-P01.csv, Rral-Storyboards-P22.csv
Nêlêmwa	19	Nêlêmwa -Storyboards-P01.csv, Nêlêmwa -Storyboards-P19.csv
Iai	19	Iai -Storyboards-P01.csv, Iai -Storyboards-P19.csv

In the individual participant files the classifier response column gives the semantic code for the classifier given for that picture:

- GEN 'general classifier'

- FOOD 'food classifier'
- CNSM 'consumable classifier'
- DRINK 'drink classifier'
- LAND 'land classifier'
- FIRE 'fire classifier'
- CARRY/BURDEN 'carry or burden classifier (for things carried)'
- CATCH 'catch classifier (game or fish caught)'
- CATCH-LINE 'catch classifier (for fish caught on a line)'
- CATCH-WEAPON 'classifier for animals caught using a weapon'
- WEAPON 'weapon classifier (for weapons or animals hunted using a weapon)'
- PLANT 'plant classifier (for things planted)'
- SUGARCANE 'sugarcane classifier'
- CHEW 'chew classifier (for chewable food)'
- ANIMAL 'domestic animal classifier'
- THOUGHT 'thought/nominalisation classifier'
- WASH 'wash classifier used for water that is used to wash with' (similar to BATHE but different etymology)
- STARCH 'starch food classifier'
- MEAT 'meat food classifier'
- VEGETABLE 'vegetable food classifier'
- CONTENTS-ROPE 'rope classifier (for things bundled up in a rope)'
- CONTENTS-BASKET 'contents of the basket classifier (for things in a basket)'
- CONTENTS-CATCH 'classifier for the result of your hunting or fishing'
- CONTENTS-TRAP 'classifier for the result of your hunting with a trap'
- BATHE 'classifier for water to wash with'
- ASS 'associative possessive preposition'
- GEN-PREP 'general possessive preposition'
- N 'No classifier used'

The Anaphoric response column had a series of codes dependent on whether the possessed noun was mentioned and in what way,

CL	No noun, just the classifier to mark the antecedent
SN	Fully specified noun that matches the target referent of the picture
HN	Hypernymic noun used to mark the antecedent noun (e.g. 'food' used to refer to 'yam')
PS	Pronominal suffix on the verb to mark the antecedent noun (not available in all languages)
PN	Pronominal free form to mark the antecedent noun
NUM	Numeral used to mark the antecedent noun (e.g., 'they took two')
ANA	A special anaphoric form used to mark the antecedent noun
DPN	A directly possessed noun
Z	Zero anaphor/gap strategy (no classifier, no noun)
N	No anaphoric device (also encoded as zero, but syntactic evidence to show no anaphor)

Data preparation for individual languages

The following data workflow shows the IPYNB Python files for each language, what files were input; a description of what they do; what files are output; and where else these output files are used.

laai:

IPYNB or R file	Input files	Output Files	what's in the output file	Next stage of analysis
laai-Anaphoric.ipynb	individual participant csv files	laai_Semantics.csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifier used)	Not used further
		laai_Anaphoric.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Anaphoric.ipynb
		laaiReg.csv	Dataframe for use in GLMM	Comparison-Anaphoric.ipynb
laai-Possessive-Constructions.ipynb	individual participant csv files	laai_Semantics_Possessive_Constructions.csv	list of each participant, the picture and the storyboard along with the different possessive constructions used (note N - indicates No possessive construction used)	laai-Constructions.ipynb
		laai_Possessive_Constructions.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Classifier-Agreement.ipynb; Comparison-Storyboard.ipynb
		laaiRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparison-Classifier-Agreement.ipynb
laai-Patterns.ipynb	individual participant csv files	laai_Patterns.csv	combined pattern for each participant/storyboard combination	Comparison-Pattern.ipynb
laai-Constructions.ipynb	laai_Semantics_Possessive_Constructions.csv	laai_constructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	Comparison-Classifier-Agreement.ipynb; Comparison-Pattern.ipynb

Lewo:

Lewo-Anaphoric.ipynb	individual participant csv files	Lewo_Semantics.csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifier used)	Not used further
		Lewo_Anaphoric.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Anaphoric.ipynb
		LewoReg.csv	Dataframe for use in GLMM	Comparison-Anaphoric.ipynb
Lewo-Possessive-Constructions	individual participant csv files	Lewo_Semantics_Possessive_Constructions	list of each participant, the picture and the storyboard along with the different possessive constructions used (note N - indicates No possessive construction used)	Lewo-Constructions.ipynb

ctions.ipnyb		nstruction s.csv		
		Lewo_Possessive_Construction s.csv	dataframe ofr 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a stroyboard;; NaN shows not enough data to say eithermatchin g or non-matching classifiers sued	Comparison-Classifer-Agreement.ipnyb; Comparison-Storyboard.ipnyb
		LewoRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparison-Classifer-Agreement.ipnyb
Lewo-Patterns.ipnyb	individual participant csv files	Lewo_Patterns.csv	combined pattern for each participant/storyboard combination	Comparison-Pattern.ipnyb
Lewo-Constructions.ipnyb	Lewo_Semantics_Possessive_Construction s.csv	Lewo_constructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	

North Ambrym (Rral):

Rral-Anaphoric.ipnyb	individual participant csv files	Rral_Semantics.csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifeir used)	Not used further
		Rral_Anaphoric.csv	dataframe ofr 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a stroyboard;; NaN shows not enough data to say eithermatchin g or non-matching classifiers sued	Comparis on-Anaphoric.ipnyb
		RralReg.csv	Dataframe for use in GLMM	Comparis on-Anaphoric.ipnyb
Rral-Possessive-Constructions.ipnyb	individual participant csv files	Rral_Semantics_Possessive_Constructions.csv	list of each participant, the picture and the storyboard along with the different possessie constructions used (note N - indicates No possessive construction used)	Rral-Construct ions.ipnyb
		Rral_Possessive_Constructions.csv	dataframe ofr 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a stroyboard;; NaN shows not enough data to say eithermatchin g or non-matching classifiers sued	Comparis on-Classifer-Agreeme nt.ipnyb
		RralRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparis on-Classifer-Agreeme nt.ipnyb
Rral-Patterns.ipnyb	individual participant csv files	Rral_Patterns.csv	combined pattern for each participant/storyboard combination	Comparis on-Pattern.i pnyb
Rral-Construct ions.ipnyb	Rral_Semantics_Possessive_Constructions.csv	Rral_const ructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	

Vatlongos:

Vatlongo s- Anaphoric.ipnyb	individual participant csv files	Vatlongos_ Semantics. csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifier used)	Not used further
		Vatlongos_ Anaphoric. csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison- Anaphoric.ipnyb
		VatlongosReg.csv	Dataframe for use in GLMM	Comparison- Anaphoric.ipnyb
Vatlongo s- Possessive- Constructions.ipnyb	individual participant csv files	Vatlongos_ Semantics_ Possessive_ Constructions.csv	list of each participant, the picture and the storyboard along with the different possessive constructions used (note N - indicates No possessive construction used)	Vatlongos- Constructions.ipnyb
		Vatlongos_ Possessive_ Constructions.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison- Classifier- Agreement.ipnyb; Comparison- Storyboard.ipnyb
		VatlongosRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparison- Classifier- Agreement.ipnyb
Vatlongo s- Patterns.ipnyb	individual participant csv files	Vatlongos_ Patterns.csv	combined pattern for each participant/storyboard combination	Comparison- Pattern.ipnyb
Vatlongo s- Constructions.ipnyb	Vatlongos_ Semantics_ Possessive_ Constructions.csv	Vatlongos_ Constructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	Comparison- Classifier- Agreement.ipnyb; Comparison- Pattern.ipnyb

Nêlêmwa:

Nelemwa- a- Anaphoric.ipnyb	individual participant csv files	Nelemwa_ Semantics. csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifier used)	Not used further
		Nelemwa_ Anaphoric. csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison- Anaphoric.ipnyb
		NelemwaReg.csv	Dataframe for use in GLMM	Comparison- Anaphoric.ipnyb
Nelemwa- a- Possessive- Constructions.ipnyb	individual participant csv files	Nelemwa_ Semantics_ Possessive_ Constructions.csv	list of each participant, the picture and the storyboard along with the different possessive constructions used (note N - indicates No possessive construction used)	Nelemwa- Constructions.ipnyb

		Nelemwa_Possessive_Constructions.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Classifer-Agreement.ipynb;; Comparison-Storyboard.ipynb
		NelemwaRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparison-Classifer-Agreement.ipynb
Nelemwa-Patterns.ipynb	individual participant csv files	Nelemwa_Patterns.csv	combined pattern for each participant/storyboard combination	Comparison-Pattern.ipynb
Nelemwa-Constructions.ipynb	Nelemwa_Semantics_Possessive_Constructions.csv	Nelemwa_constructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	Comparison-Classifer-Agreement.ipynb; Comparison-Pattern.ipynb

Merei:

Merei-Anaphoric.ipynb	individual participant csv files	Merei_Semantics.csv	list of each participant, the picture and the storyboard along with the classifiers used anaphorically (note N - indicates No classifier used)	Not used further
		Merei_Anaphoric.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching classifiers used within a storyboard; 0 = non-matching classifiers used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Anaphoric.ipynb
		MereiReg.csv	Dataframe for use in GLMM	Comparison-Anaphoric.ipynb
Merei-Possessive-Constructions.ipynb	individual participant csv files	Merei_Semantics_Possessive_Constructions.csv	list of each participant, the picture and the storyboard along with the different possessive constructions used (note N - indicates No possessive construction used)	Merei-Constructions.ipynb
		Merei_Possessive_Constructions.csv	dataframe of 1/0 and NaN for each participant/storyboard. 1 = matching possessive constructions used within a storyboard; 0 = non-matching constructions used within a storyboard;; NaN shows not enough data to say either matching or non-matching classifiers used	Comparison-Classifer-Agreement.ipynb; Comparison-Storyboard.ipynb
		MereiRegPossessiveConstructions.csv	Dataframe for use in GLMM	Comparison-Classifer-Agreement.ipynb
Merei-Patterns.ipynb	individual participant csv files	Merei_Patterns.csv	combined pattern for each participant/storyboard combination	Comparison-Pattern.ipynb
Merei-Constructions.ipynb	Merei_Semantics_Possessive_Constructions.csv	Merei_constructions.csv	calculates number of different possessive constructions a participant used across the eight storyboards	Comparison-Classifer-Agreement.ipynb; Comparison-Pattern.ipynb

Combining individual language data

The following data files combine the prepared data files described above and prepare output csv files for use in R for data analysis.

Python file for anaphoric uses of classifiers:

IPYNB or R file	Input files	Output Files	what's in the output file	Next stage of analysis
Comparison-Anaphoric.ipnyb	Merei_Anaphoric.csv, Lewo_Anaphoric.csv, Vatlongos_Anaphoric.csv, Rral_Anaphoric.csv, laai_Anaphoric.csv, Nelemwa_Anaphoric.csv	Combined_Anaphoric.csv	combined data frame of each language for matching/non-matching anaphoric use of classifiers across all storyboards, along with percentages, and basic socio information (age, sex and education)	Storyboards_Analysis.Rmd
	MereiReg.csv, LewoReg.csv, RralReg.csv, VatlongosReg.csv, NelemwaReg.csv, laaiReg.csv	RegCombAna.csv	Combined data frame for use in GLMM	Storyboards_Analysis.Rmd

Python file for comparing all different possessive constructions:

IPYNB or R file	Input files	Output Files	what's in the output file	Next stage of analysis
Comparison-Possessive-Constructions.ipnyb	Merei_Possessive_Constructions.csv, Merei_constructions.csv, Lewo_Possessive_Constructions.csv, Lewo_constructions.csv, Vatlongos_Possessive_Constructions.csv, Vatlongos_constructions.csv, Rral_Possessive_Constructions.csv, Rral_constructions.csv, laai_Possessive_Constructions.csv, laai_constructions.csv,	Combined_Classifier_Agreement.csv	combined data frame of each language for matching/non-matching possessive construction across all storyboards, along with percentages, and basic socio information (age, sex and education), and overall number of possessive constructions used per participant across all storyboards	Storyboards_Analysis.Rmd

	Nelemwa_Possessive_Constructions.csv, Nelemwa_constructions.csv			
	MereiRegPossessiveConstructions.csv, LewoRegPossessiveConstructions.csv, RralRegPossessiveConstructions.csv, VatlongosRegPossessiveConstructions.csv, NelemwaRegPossessiveConstructions.csv, laaiRegPossessiveConstructions.csv	RegCombPossessiveConstructions.csv	Combined data frame for use in GLMM	Storyboards_Analysis.Rmd

Python file for different abstract patterns:

IPYNB or R file	Input files	Output Files	what's in the output file	Next stage of analysis
Comparison - Pattern.ipynb	Merei_Patterns.csv	Merei_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
	Lewo_Patterns.csv	Lewo_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
	Vatlongos_Patterns.csv	atlongos_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
	Rral_Patterns.csv	Rral_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
	laai_Patterns.csv	laai_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
	Nelemwa_Patterns.csv	Nelemwa_paired_t_test.csv	counting total unique number of matching and non-matching patterns per individual participant to use for paired t-tests	Storyboards_Analysis.Rmd
		Pattern_Sample.csv	creating percentage of each pattern for each language	Storyboards_Analysis.Rmd
		Counts2.csv	counts the overall number of patterns used by each participant by each language for use in the ANOVA	Storyboards_Analysis.Rmd
	Merei_constructions.csv	PatternConstructions.csv	creating one df for matching and non-matching unique patterns, total and number of constructions	Storyboards_Analysis.Rmd
	Lewo_constructions.csv			
	Vatlongos_constructions.csv			
	Rral_constructions.csv			
	laai_constructions.csv			

	Nelemwa_constructions.csv			
--	---------------------------	--	--	--

Python file for comparing across storyboards:

IPYNB or R file	Input files	Output Files	what's in the output file	Next stage of analysis
Comparison-Storyboard.ipnyb	Merei_Possessive_Constructions.csv	Merei_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
	Lewo_Possessive_Constructions.csv	Lewo_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
	Vatlongos_Possessive_Constructions.csv	Vatlongos_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
	Rral_Possessive_Constructions.csv	Rral_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
	Iaai_Possessive_Constructions.csv	Iaai_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
	Nelemwa_Possessive_Constructions.csv	Nelemwa_Possessive_Constructions_Chisq.csv	use of possessive constructions, but with all participant/storyboard matching/non-matching answers for use with Chisq	Storyboards_Analysis.Rmd
		Combined_Storyboard_Possessive_Constructions.csv	creating a dataframe for the matching/non-matching possessive constructions comparing across storyboards and languages using the percentage data for use in R	Storyboards_Analysis.Rmd

R Statistical Analysis

The Storyboards_Analysis.Rmd file has all the statistical analysis for the submitted paper. Each chunk corresponds to a particular section of the analysis (section 4) in the paper.