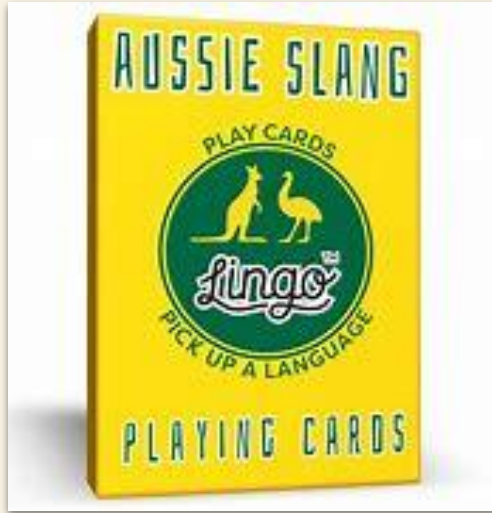


# Sharing the Australian Slang Survey Data collection

Rosanna Smith  
Simon Musgrave



# The project



- The data was collected as part of the project 'Metaphors and Identities in the Australian Vernacular' supported by the Australian Research Council (SR200200350).
- The Chief Investigators for the project were:
  - Professor Kate Burridge
  - Dr Howard Manns
  - Dr Simon Musgrave
  - Emeritus Professor Keith Allen
- Research Associate: Dr Isabelle Burke
- PhD Student: Dylan Hughes
- All attached to Monash University.

# The survey

- Aimed to access knowledge:
  - Of users of Australian English
  - About Australian slang
- Collected basic demographic information:
  - Gender
  - Age (in brackets)
  - Place of residence
  - Childhood location (if different from current place of residence)
- Body of survey is responses to a series of prompts
- Survey delivered via Qualtrics
- Responses collected July and August 2021
  - More than 4000 people visited the survey
  - Almost 3000 contributed responses



# Prompts

- Participants were asked to provide information in response to 13 prompts
- Response categories:
  - Word or expression:
  - Meaning of word/expression:
  - Example of word/expression in use:
  - When and where have you heard this word/expression used?
  - Any further comments:



# Prompts

1. Something very good?
2. Something very bad?
3. Someone who is stupid?
4. An attractive person of any gender?
5. An attractive female?
6. An attractive male?
7. An unattractive person?
8. Someone who is arrogant?
9. Nonsense?
10. Alcohol?
11. Being intoxicated?
12. Someone who does not do their fair share?
13. Any part of the body?

Open question:

Lastly, what other words or expressions do you think of as being characteristically Australian?

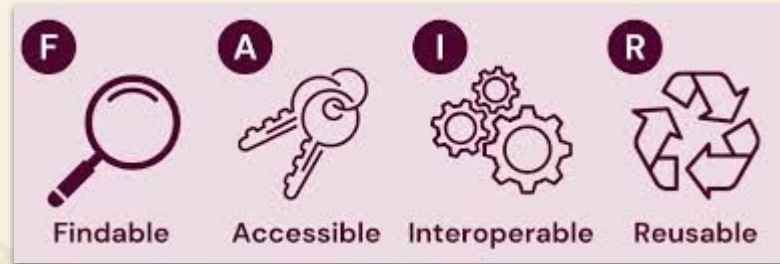
# What the responses look like

24	Start Date	End Date	Response IP Address	Progress	Duration (Finished)	Recorded Date	Response Recipient	Recipi	Recip	Ext	Location L	Location L	Distributi	User Lang	Is your agi	In what ra	What i	What genc	My gender	Where do	Did you gr	Where did if	If you wou
25	20/07/2021 4:04	20/07/2021 4:26	IP Address: 121.45.136	100	1319	TRUE	20/07/2021 4:26 R_3j0wrfek7k9iVO3				-27.4732	153.02151	anonymoi	EN	Yes	40 - 60	Woman			4573	No	Sydney NS	robertsona
26	20/07/2021 4:21	20/07/2021 4:35	IP Address: 120.149.81	100	889	TRUE	20/07/2021 4:35 R_2y8gILYzDBaSe2N				-31.9474	115.86480	anonymoi	EN	Yes	Over 60	Woman			No	Highgate	harleyfam	
27	20/07/2021 3:37	20/07/2021 4:41	IP Address: 124.170.45	100	3841	TRUE	20/07/2021 4:41 R_10CvZlGtuabviV0				-31.9474	115.86480	anonymoi	EN	Yes	40 - 60	Woman			6016	No	6100	
28	20/07/2021 4:08	20/07/2021 4:47	IP Address: 49.199.49.	100	2339	TRUE	20/07/2021 4:47 R_1QuRC1uLCQOGXFW				-37.8109	144.95440	anonymoi	EN	Yes	40 - 60	Woman			3073	Yes		pjanemor
29	20/07/2021 5:02	20/07/2021 5:03	IP Address: 115.189.91	100	60	TRUE	20/07/2021 5:03 R_2yq1pb6pBITV0aF				-36.8506	174.76791	anonymoi	EN	Yes	40 - 60	Woman						
30	20/07/2021 4:35	20/07/2021 5:18	IP Address: 1.136.22.7	100	2570	TRUE	20/07/2021 5:18 R_27P772ueSBI00s3				-37.8159	144.96688	anonymoi	EN	Yes	19 - 40	Man			3070	No	Southeast L	cheonph

24	Do you kn	What is th	Can you provide an example of the word or expressi	When and where have you heard this word or expression be	Is there anything else you would like to tell us about this word c
25	ann@hotmail.com				
26	Yes	Bonza	Thatâ€™s a bonza idea	A few years ago now	
27	Yes	Bonza	Thatâ€™s bonza mate	In my childhood	
28	Yes	Cracker	Thatâ€™s an absolute cracker	Use it myself	
29					
30	Yes	Bloody rip	Thatâ€™s bloody ripper	Itâ€™s currently on a PBS radio ad in Melbourne	
31	Yes	Bonza	That ute is bonza		
32	Yes	Awesome	Thatâ€™s awesome	Anytime something good happens	
33	Yes	beaut	Mate, that car is a real beaut	Mainly by older whiter speakers	Seems a bit old fashioned, I would rarely use this
34	Yes	Grouse	That's grouse!		

# Sharing the data - FAIRification

- Previous slide already shows some problems for sharing this data
  - Potentially identifying information (location when responding, email in some cases)
  - Encoding problems ('Often used with â€˜bloodyâ€™ before it (bloody ripper)')
  - Unwieldy data structure - we showed responses to one prompt, all responses for a participant (up to 14) are a single row in the table





# Data transformation using Jupyter Notebooks

- Input: Original survey data consisting of 121 columns
- Required outputs:
  - 14 CSV files per prompt
  - De-identified version of the original data
- [Slang\\_Data\\_Processing.ipynb](#)

## Slang Data Processing

This notebook is used for processing the file *Australian Slang v2\_September 5, 2021.csv* which is the original data from the Australian Slang Survey Data collection.

The notebook will create two outputs:

1. A CSV folder with 14 CSV files - one for each of the slang prompts given to the participants. Non-responses have been removed, participant IDs added, and minor character encoding and formatting applied.
2. A de-identified version of the original survey data, where seven columns have been removed.

## Import libraries

```
In [ ]: import pandas as pd
import re
import os
```

## Load Australian Slang v2\_September 5, 2021.csv as a dataframe

```
In [ ]: # Load Australian Slang v2_September 5, 2021.csv
df = pd.read_csv('Australian Slang v2_September 5, 2021.csv')

# Show the first few rows
df.head()
```



# Identifying information

Columns removed:

- IPAddress
- RecipientLastName
- RecipientFirstName
- RecipientEmail
- LocationLatitude
- LocationLongitude
- If you would be willing to answer further questions, please provide an emailaddress.

## Remove identifying columns

Columns removed:

- 3: IPAddress
- 9: RecipientLastName
- 10: RecipientFirstName
- 11: RecipientEmail
- 13: LocationLatitude
- 14: LocationLongitude
- 25: If you would be willing to answer further questions, please provide an emailaddress.

```
In [ ]: data = orig.drop(orig.columns[[3, 9, 10, 11, 13, 14, 25]], axis=1)
data.head()
```

# Prompt separation

- Dataframe split according to each prompt
- Columns kept for CSVs:
  - Response ID
  - Do you know any typically Australian word or expression for [prompt]?
  - What is the word or expression?
  - What does this word or expression mean? (Note: Where included in the survey)
  - Can you provide an example of the word or expression in use?
  - When and where have you heard this word or expression being used?
  - Is there anything else you would like to tell us about this word or expression?
- Prompt column added

## Separate data into prompts

Selecting only the columns:

- Response ID
- Do you know any typically Australian word or expression for [prompt]?
- What is the word or expression?
- What does this word or expression mean? (Note: Where included in the survey)
- Can you provide an example of the word or expression in use?
- When and where have you heard this word or expression being used?
- Is there anything else you would like to tell us about this word or expression?

```
In [ ]: verygood = df.iloc[:, [8, 26, 27, 28, 29, 30]]
verybad = df.iloc[:, [8, 31, 32, 33, 34, 35]]
stupid = df.iloc[:, [8, 36, 37, 38, 39, 40]]
attractiveperson = df.iloc[:, [8, 41, 42, 43, 44, 45]]
attractivefemale = df.iloc[:, [8, 46, 47, 48, 49, 50]]
attractivemale = df.iloc[:, [8, 51, 52, 53, 54, 55]]
unattractiveperson = df.iloc[:, [8, 56, 57, 58, 59, 60]]
arrogant = df.iloc[:, [8, 61, 62, 63, 64, 65]]
nonsense = df.iloc[:, [8, 66, 67, 68, 69, 70]]
alcohol = df.iloc[:, [8, 71, 72, 73, 74, 75]]
intoxicated = df.iloc[:, [8, 76, 77, 78, 79, 80]]
doesntdoairshare = df.iloc[:, [8, 81, 82, 83, 84, 85]]
bodypart = df.iloc[:, [8, 86, 87, 88, 89, 90, 91]] # From here onwards there are 7 columns
freechoice_a = df.iloc[:, [8, 92, 93, 94, 95, 96, 97]]
freechoice_b = df.iloc[:, [8, 98, 99, 100, 101, 102, 103]]
freechoice_c = df.iloc[:, [8, 104, 105, 106, 107, 108, 109]]
freechoice_d = df.iloc[:, [8, 110, 111, 112, 113, 114, 115]]
freechoice_e = df.iloc[:, [8, 116, 117, 118, 119, 120, 121]]
```

## Insert 'prompt' columns

```
In [ ]: verygood.insert(loc=1, column='prompt', value='Very Good')
verybad.insert(loc=1, column='prompt', value='Very Bad')
stupid.insert(loc=1, column='prompt', value='Stupid')
attractiveperson.insert(loc=1, column='prompt', value='Attractive Person')
attractivefemale.insert(loc=1, column='prompt', value='Attractive Female')
attractivemale.insert(loc=1, column='prompt', value='Attractive Male')
unattractiveperson.insert(loc=1, column='prompt', value='Unattractive Person')
arrogant.insert(loc=1, column='prompt', value='Arrogant')
nonsense.insert(loc=1, column='prompt', value='Nonsense')
alcohol.insert(loc=1, column='prompt', value='Alcohol')
intoxicated.insert(loc=1, column='prompt', value='Intoxicated')
doesntdoairshare.insert(loc=1, column='prompt', value='Doesn't Do Fair Share')
bodypart.insert(loc=1, column='prompt', value='Body Part')
freechoice_a.insert(loc=1, column='prompt', value='(Free Choice)')
freechoice_b.insert(loc=1, column='prompt', value='(Free Choice)')
freechoice_c.insert(loc=1, column='prompt', value='(Free Choice)')
freechoice_d.insert(loc=1, column='prompt', value='(Free Choice)')
freechoice_e.insert(loc=1, column='prompt', value='(Free Choice)')
```

# Handling prompts 13 and 14

Prompts for Body Part and (Free Choice) had an additional column 'meaning' for participants to describe the word or expression.

Additional blank 'meaning' column added to other prompts so data is structured identically.

Item	Meaning
What is the word or expression?	What does this word or expression mean?
God stone the crows	An exclamation of frustration or exasperation
Drongo	Idiot
Cobber	A friend
Snag	Sausage
grouse	very good
Ta	Thank you
Yobbo	A useless bogan
Cobba	Mate

Insert empty 'meaning' columns for prompts missing this

```
In [ ]: verygood.insert(loc=4, column='meaning', value='')
        verybad.insert(loc=4, column='meaning', value='')
        stupid.insert(loc=4, column='meaning', value='')
        attractiveperson.insert(loc=4, column='meaning', value='')
        attractivefemale.insert(loc=4, column='meaning', value='')
        attractivemale.insert(loc=4, column='meaning', value='')
        unattractiveperson.insert(loc=4, column='meaning', value='')
        arrogant.insert(loc=4, column='meaning', value='')
        nonsense.insert(loc=4, column='meaning', value='')
        alcohol.insert(loc=4, column='meaning', value='')
        intoxicated.insert(loc=4, column='meaning', value='')
        doesntdoairshare.insert(loc=4, column='meaning', value='')
```

# Non-responses

- Blank prompt rows removed from data
- Updated structure preserves the other responses provided by that participant

Do you know any typically Australian word or expression for something very good?	What is the word or expression?
Yes	Cracker
Yes	Bloody ripper
Yes	Bonza
Yes	Awesome
Yes	beaut
Yes	Grouse
No	
Yes	Beaut, top notch
Yes	Grouse, ace, excellent, cool, rad, sweet, awesome, bonza

## Remove rows where 'slang' is empty

```
In [ ]: combined_df = combined_df[combined_df['slang'].notna() & (combined_df['slang'].str.strip())]
combined_df = combined_df.drop(columns='known')
combined_df.head(15)
```

# Formatting and character encoding

- Leading and trailing spaces removed
- Quotation mark standardisation
- Minor character encoding and formatting corrections

"You're a drongo", "You stupid Galah"  
"you got the job?! That's grouse!"  
It may also simply mean 'lazy' in some contexts.  
You're such a dork, Bronson.  
Don't be a Drongo  
Bring a bottle of plonk  
wine o'clock  
Fair dinkum ?  
Used by a previous generation  
he is just a bludger, leaves all the work for someone else

"You're a drongo", "You stupid Galah"  
"you got the job?! That's grouse!"  
It may also simply mean 'lazy' in some contexts.  
You're such a dork, Bronson.  
Don't be a Drongo  
Bring a bottle of plonk  
wine o'clock  
Fair dinkum?  
Used by a previous generation  
he is just a bludger, leaves all the work for someone else

# Mapping the participant metadata

- Participant ID maps the responses to the participant metadata
- Prompt data exported to 14 CSVs
- Geomapping data for postal areas added to participant metadata based on their postcodes provided

```
{
  "@id": "#pl3000",
  "@type": "Place",
  "name": "Postcode 3000",
  "geo": {
    "id": "#loc3000"
  }
},
{
  "@id": "#loc3000",
  "@type": "Geometry",
  "name": "Geometry 3000",
  "geo:asWKT": "POLYGON ((144.96138866600006 -37.82050130699997, 144.96138866600006 -37.82050130699997, 144.96138866600006 -37.82050130699997, 144.96138866600006 -37.82050130699997))"
```

Melbourne Postcode  
Polygon Metadata

participantID	responseID	prompt	slang	meaning	exampleUsage	whenWhereUsed	additionalNotes
A0032	R_2bJZCteOdKLHfqv	(Free Choice)	Arvo	Afternoon	'Come over this arvo'	In pretty much any context - even in a ...	
A0033	R_2CTzqrC7w1fQrC	(Free Choice)	Cobber	friend - usually a man	"Hey cobber" or "this old cobber came over" (latter	Have heard friends say it, hear older men say	
A0034	R_24pXXdrFmVGIWSI	(Free Choice)	yeah nah	yes I understand you, no I don't agree	yeah nah	everywhere	
A0035	R_2VPAgu108j33uuh	(Free Choice)	G'day	Hello	G'day stranger! Haven't seen you for ages!	I use it all the time. It's a classic!	
A0036	R_3hgYG7y4AJ61SE	(Free Choice)	sook	soppy, clingy, loveable	my cat follows me everywhere, he's such a sook		
A0037	R_1igFsZGdFjHsDmr	(Free Choice)	Mad as a cut snake	Very angry	He's so angry; he's mad as a cut snake	In situations of conflict	Someone is so mad they could lose control ...
A0038	R_20MYGo3v8pghrl2	(Free Choice)	various hyponyms				question is too vague
A0039	R_WfB9hSVjmG63o4h	(Free Choice)	chock-a-block, chockers	full, packed, jammed due to number of ...	The freeway's chockers at the moment. I might just	Often used in traffic reports	



# Data in LDaCA Portal

[data.ldaca.edu.au](http://data.ldaca.edu.au)



## Australian Slang Survey Data

Name	Australian Slang Survey Data
Description	This dataset consists of responses to an online survey which elicited information about Australian slang. Respondents were asked to provide slang expressions equivalent to a variety of prompts (e.g. 'Something very bad') along with additional information such as who they might expect to use the expression given. The data was collected as part of the project 'Metaphors and Identities in the Australian Vernacular' supported by the Australian Research Council (SR200200350).
Date Published	2025-10-28
ID	arcp://name,hdl10.26180-30102115
Author	<a href="#">Kate Burridge</a> <a href="#">Howard Manns</a> <a href="#">Simon Musgrave</a> <a href="#">Isabelle Burke</a> <a href="#">Dylan Hughes</a> <a href="#">Keith Allan</a>
Related Works	<a href="#">Burke, I., &amp; Manns, H. (2024). Australia's idiomatic expressions: "Speaking the culture" to manage social relations. Australian Journal of Linguistics, 44(4), 365-389. https://doi.org/10.1080/07268602.2024.2449060</a> <a href="#">Burridge, K. (2024). "Survival of the fittest" - the evolution of language. Australian Journal of Linguistics, 44(4), 340-364. https://doi.org/10.1080/07268602.2024.2425358</a> <a href="https://w3id.org/ldac/profile#Collection">https://w3id.org/ldac/profile#Collection</a>
Conforms To	<a href="#">https://w3id.org/ldac/profile#Collection</a>
Credit Text	Burridge, Kate; Manns, Howard; Musgrave, Simon; Burke, Isabelle; Hughes, Dylan; Allen, Keith (2025). Australian Slang Survey Data. Monash University. Dataset. <a href="https://doi.org/10.26180/30102115.v1">https://doi.org/10.26180/30102115.v1</a>
In Language	<a href="#">Australian English</a>
Communication Mode	<a href="https://w3id.org/ldac/terms#WrittenLanguage">https://w3id.org/ldac/terms#WrittenLanguage</a>
DOI	<a href="https://doi.org/10.26180/30102115">https://doi.org/10.26180/30102115</a>
Has Collection Protocol	Slang Data Provenance
Metadata Is Public	true
Subject Language	<a href="#">Australian English</a>
Publisher	<a href="#">Monash University</a>

Objects in Collection: 15

[Collections](#) [Notebooks](#) [Browse](#) [Login](#) [Help](#)

### Access

You are free to:  
Share — copy and redistribute the material in any medium or format  
The licensor cannot revoke these freedoms as long as you follow the license...

[Read more](#)

[Attribution-NonCommercial-NoDerivatives 4.0 International](#)

Public Metadata and Text is Searchable

### Content

Language  
Australian English  
File Formats  
text/csv

### Downloads

Australian Slang Survey Data.zip  
Files: 18, Size: 218.06 MB  
[Attribution-NonCommercial-NoDerivatives 4.0 International](#)



[Show All Related Downloads](#)

### Retrieve Metadata

[Download metadata](#)  
[Open metadata in a new window](#)

Metadata licensed as:  
[Attribution 4.0 International \(CC BY 4.0\)](#)



[Collections](#)

[← Item12 - Doesn't Do Fair Share](#) > [Item12-DoesntDoFairShare.csv](#)

participantID	responseID	prompt	slang	meaning	exampleUsage	wh
A0001	R_3MarUAEXgFBBuGf	Doesn't Do Fair Share	bludger			
A0002	R_3pmQPCnOfdyR2No	Doesn't Do Fair Share	Bludger		That bastard is a total bludger	I us
A0003	R_3qVjXWdknzHy4j5	Doesn't Do Fair Share	a slacker		He has always been a slacker	eve
A0004	R_3rGNrQUp49FGe48	Doesn't Do Fair Share	Bludger		He won't be any help beca use he is just a bludger	
A0005	R_1mK8LtcwTvh08W	Doesn't Do Fair Share	Bludger		He is such a bludger	In s
A0006	R_reDFubAmVVOPIl	Doesn't Do Fair Share	Bludger		A dole bludger	TV
A0007	R_Tba7dCWNOzI4iHL	Doesn't Do Fair Share	bludger		dole-bludgers	me d w on ear ay n.
A0008	R_3QJeHx8WNfrL3Q	Doesn't Do Fair Share	Bludger		He's a bludger	
A0009	R_xgDzRk6iMsmtdR7	Doesn't Do Fair Share	Shirk		He shirked his responsibilities	



# Future work

Improve search capabilities and display in the LDaCA portal.

Add an additional download for a combined prompts CSV.

Create a Jupyter Notebook that allows users to explore the collection:

- Filter by a word or expression and view its related demographics
- View most common responses to a prompt and any associated geographical patterns
- Geomapping a given word or expression using the postcode data provided by the participants

## Item01-VeryGood.csv

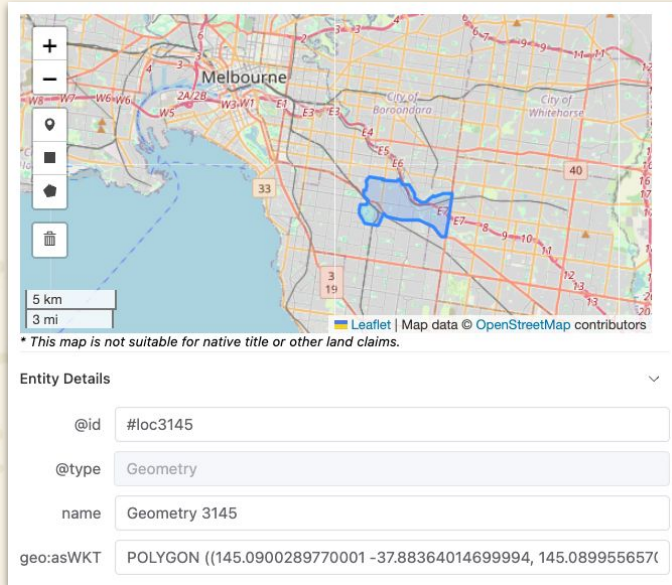
Type: File Idac:DerivedMaterial csvw:Table

Member of: [Australian Slang Survey Data](#)

...A0633,R\_3dLCx32wc2zMsFL,Very Good,Ripper,,Gee we had a ripper time at the 'G', "Ripper in general sense.G'day, ripper day mate....

👤 Relevance Score: 11.567043

[See more](#)





**Language Data Commons of Australia**



**ldaca.edu.au**



**ldaca@uq.edu.au**



**@LDaCA\_Project**



**@ldaca**



Australian Research Data Commons



Australian Government



National Research  
Infrastructure for Australia

ARDC is  
enabled  
by NCRIS

The **Language Data Commons of Australia (LDaCA)** receives investment from the **Australian Research Data Commons (ARDC)** through its **HASS and Indigenous Research Data Commons**.

The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).



**Stay in the Loop**

join our newsletter to  
receive quarterly updates