LDaCA

Language Data Commons of Australia

2025

YEAR IN REVIEW



FOREWORD

2025 was a productive year for the Language Data Commons of Australia (LDaCA). We expanded collections, strengthened our tools and supported a wide range of research using LDaCA resources.

This year, LDaCA collections and tools were used in studies on Australian slang, parliamentary language, online discourse and even featured in a number of news stories. We delivered well-attended workshops and webinars and delivered another iteration of the Graduate Digital Research Fellowship program. We also collaborated on cross-institutional projects, including the Precarious Victorian Oral History Collections Project and the development of person-centred digital collections such as *Arne ingkerreke apurtelhe-ileme*, the Veronica Dobson website. Our technology and research teams also advanced RO-Crate interoperability, released new tools and deepened partnerships across the sector.

In 2026, we are excited to continue improving LDaCA platforms and resources. We are looking forward to developing workflows for interactive deposits, broadening the range of language data available and releasing new analytical notebooks and tools. We are also committed to sharing LDaCA's approaches more widely across disciplines and further strengthening our governance frameworks to support ethical, sustainable data stewardship.

We extend our thanks to the LDaCA team, our partners and our Steering Committee for their steady support throughout the year. Their expertise and collaboration have been central to progressing our technical, governance and engagement work. We are deeply grateful for the commitment each person brings to the project.

Michael Haugh
Project Director

Robert McLellan
Senior Program Manager



Image Robert McLellan and Michael Haugh. Photo by David Hannah, supplied by the ARDC.

> **LDaCA continues to grow as a sustainable, ethical and practical infrastructure for researchers. Our progress this year reflects the power of collaboration across teams, partners and communities.**

BY THE NUMBERS

35

conference presentations

View the full list >

16

workshops, training and webinars

with attendance of over **400** people

View the full list >

14

academic papers

View the full list >

04

new analytics tools released

Find out more >

04

datasets added to the data portal

Find out more >

13

GitHub repositories created

Access them on GitHub >

17

online articles and blogs

View our blog >

SUPPORTING DATA DISCOVERY

Demonstrating LDaCA's impact through use of language data



"More tucker than you could poke a stick at!" research paper

In January, a journal article from PhD candidate Madeleine Clews (University of Western Australia), published in the *Australian Journal of Linguistics*, showcased the impact of LDaCA collections for researchers. The article, '"More tucker than you could poke a stick at": The bicultural journey of an enduring Australianism', explored the development, use and longevity of 'tucker' in Australian speech, making comparisons between mainstream Australian English and Aboriginal English. Clews cited the Corpus of Oz Early English (COOEE), AustLit and the Australian Radio Talkback corpora, accessed through our data portal, to support her argument.

Read the article >

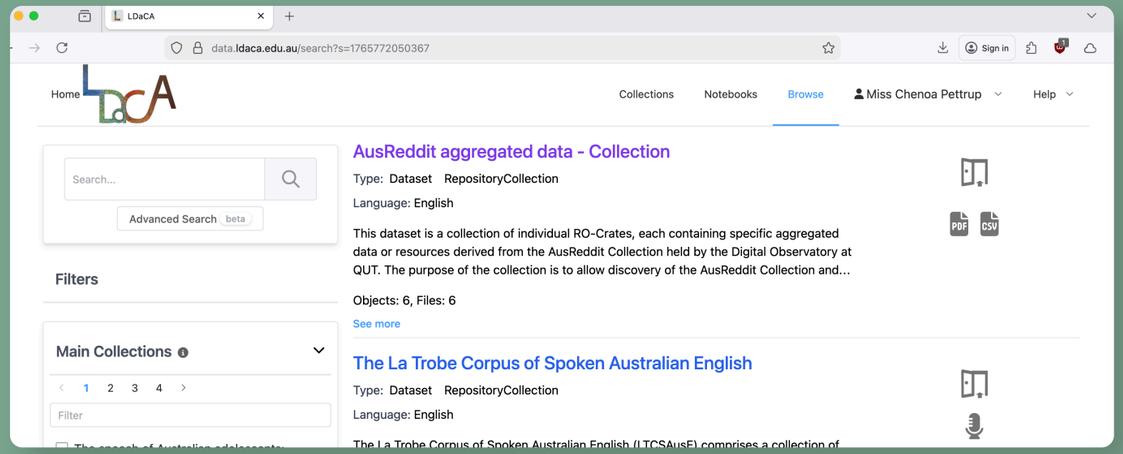


Australian slang as a literary genre

In February, Research Support and Training Lead Simon Musgrave published a journal article titled 'Australian slang as a literary genre' in the *Australian Journal of Linguistics*. Simon used four collections from the LDaCA Data Portal — the Australian Corpus of English, the Australian component of the International Corpus of English (ICE), Australian Radio Talkback and Braided Channels — to strengthen his argument that there is a tradition of Australian writers inventing expressions which they treat as being part of Australian slang. Simon also published a **blog post** discussing how vital access to these data collections was for enriching his research.

Read the article >

Highlights
new datasets on the LDaCA Data Portal



Australian Slang Survey Data
This dataset consists of responses to an online survey which elicited information about Australian slang. Respondents were asked to provide slang expressions in response to a variety of prompts.

More about the data set >

Australian Twittersphere aggregated data
This dataset is a collection of resources derived from the Australian Twittersphere Collection held by the Digital Observatory at QUT. Content examples include Australian Twittersphere hashtags, N-grams and more.

More about the data set >

Sydney Speaks
A compilation of three sub-corpora of Australian English, made up of sociolinguistic interviews and oral histories. Recordings are from a total of 260 speakers, born from the 1890s to the 1990s, recorded in the 1970s-1980s and 2010s-2020s.

More about the data set>

AusReddit aggregated data
This dataset is a collection of resources derived from the AusReddit Collection held by the Digital Observatory at QUT. Content examples include AusReddit topics, daily emotions and more.

More about the data set>

TACKLING DATA CHALLENGES

Addressing practical barriers to working with language data

“How is that unparliamentary?”

In June, Research Analytics Lead Sam Hames, Michael Haugh and Simon Musgrave published a journal article in *Lingua* titled “‘How is that unparliamentary?”: The metapragmatics of ‘unparliamentary’ language in the Australian Federal Parliament’. The article dove into data from the Australian Federal Hansard to analyse occurrences of language deemed ‘unparliamentary’. Sam, Michael and Simon drew on computational methods to explore how the boundaries of offensive, objectionable or disorderly language in parliamentary discourse are established, maintained, contested and changed over time.

In a **blog post**, Simon discussed the challenges of working with this unwieldy data source. In the post, he unpacked how they approached issues of access and scale to use this data source effectively. Sam, Michael and Simon’s work with the Australian Federal Hansard later led to a series of workshops introducing HASS researchers to the rich possibilities of working with Hansard.

Read the journal article >



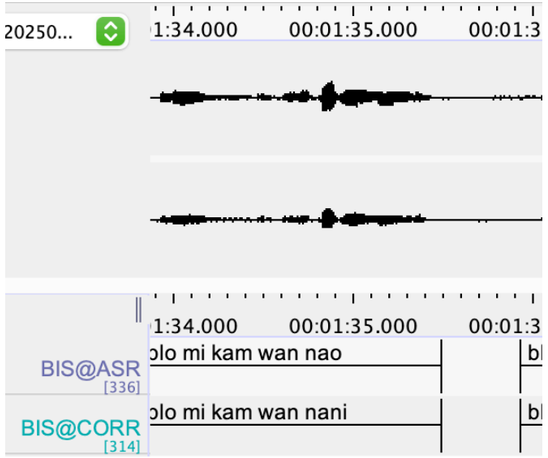
Image Simon Musgrave and Sam Hames deliver a workshop on exploring Hansard for HASS researchers.



Precarious Oral Histories

LDaCA researchers at the Australian National University launched the Identifying Precarious Victorian Oral History Collections project this year, led by Research Project Manager Anisa Puri. Oral history collections hold invaluable insights into our historical, cultural and linguistic past, yet little is known about precarious collections that are at risk of being lost. This project took an important first step by identifying and gathering metadata about such collections in Victoria. The project identified 189 collections. The collections were created between the 1960s and 2020s, spanned 29 subject areas and included a total of 9,136 interviews with an average interview length of 1.11 hours. The team produced a final report sharing these findings, as well as an online database listing the collections.

View the report > | View the database >



Transcribing and translating Bislama

Two PhD students, Raphaël Merx and Aso Mahmoudi, worked with Chief Investigator (CI) Nick Thieberger to build systems for converting Bislama (Vanuatu) audio recordings to text and then translating them into English.

Read the blog post >

BUILDING & USING TECHNOLOGIES & TOOLS

Developing tools and enabling their use in language research

Quotation Tool in use

The ATAP Quotation Tool continues to support linguistic research through the analysis of reported and quoted speech. This year, it featured in two studies examining sentiment and representation in media texts.

CI Monika Bednarek, with colleague Maite Taboada (Simon Fraser University), used the tool to compare sentiment in quoted and non-quoted news discourse in their article '**Attitude in Reported and Non-reported News: A Critique of Sentiment Analysis in Corpus Pragmatics.**' The study showed how sentiment analysis can support exploration, but should be paired with qualitative analysis.

The tool also supported University of Sydney student Melissa Kemble's PhD research. Her thesis, '**A corpus-based critical discourse analysis of gender bias and evaluation in Australian Rules Football and Rugby League sports news discourse**', analysed gender stereotypes in Australian sports reporting by examining who is quoted and how emotion is attributed to female and male athletes.

Access the Quotation Tool on Github >

AROCAPI API release

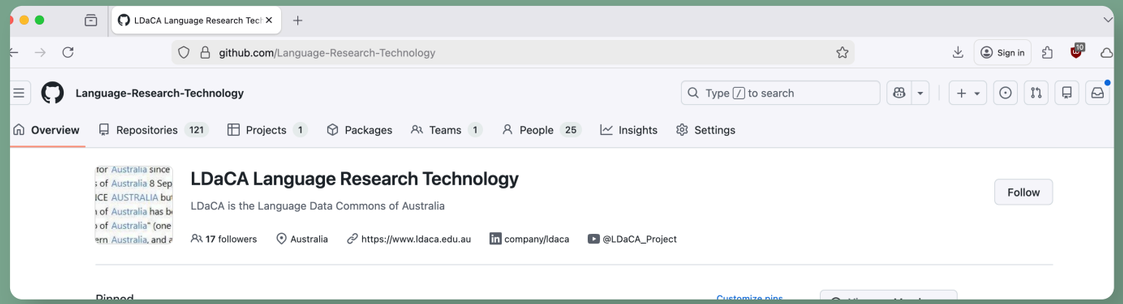
2025 saw the development of exciting work for LDaCA and its partners. PARADISEC colleague John Ferlito developed a simpler, shared API that can support both PARADISEC and LDaCA, with potential application across other repositories, including the **Nyingarn Repository**.

This new RO-Crate API, known as AROCAPI, is a generic API for collections described using RO-Crates. LDaCA has worked closely with John throughout its development, including building a new Oni stack based on the API. AROCAPI is designed to extend data portals and provide a common foundation for RO-Crate-based infrastructure. It works out of the box with standard portal interfaces, while remaining flexible enough to support different audiences, domains and authentication models.

PARADISEC has implemented the RO-Crate API in its admin tool, with the new Oni Portal set to replace the frontend in early 2026. It currently hosts 18,400 hours of audio in over 1,400 languages and continues to grow.

Access AROCAPI on Github >

Highlights
new & updated tools



Crate-O HTML Lite Display

Crate-O HTML Lite Display is a tool to create a static RO-Crate preview HTML file. It was used in the development of **Ayeye Thipe-akerte**, a a website featuring Eastern Arrernte language information about birds, with original artwork from Eastern Arrernte artist and writer Therese Ryder. Originally one in a series of bird apps produced by Batchelor Institute, it is the first to be re-released in a format based on principles of long-term safekeeping.

CADRE integration for Data Portal access

The LDaCA Data Portal's authentication process was updated to incorporate CADRE, a wrapper for REMS that improves the user experience of the authentication procedure.

RO-Crate tabulator

Sydney Informatics Hub developed the **RO-Crate Tabulator**, a Python library to turn an RO-Crate into tabular formats.

COOEE Notebook

The **COOEE Notebook** is a Jupyter notebook for viewing stratified topics in the COOEE collection.

RO-Crate MASP

The **RO-Crate Machine Actionable Schemas and Profiles** is a prototype and proof of concept for a process implementing Schemas and Profiles in RO-Crates.

Analytics tools

Four new analytics tools were released including the **Web Archive Manager (WAM)**, **Word Transcript Tabulator**, **Text Analytics Tools Web Application** and the **Image Dataset Explorer**.

Lameta update

Lameta is a tool for creating collection metadata in the field. Updates in 2025 include output to RO-Crate to fit with LDACA's technical recommendations.

Nyingarn maintenance

Nyingarn is a secure workspace for creating textual versions of manuscripts. Led by CI Nick Thieberger, it has 1,466 items and 360 users. LDACA supported ongoing maintenance work on Nyingarn in 2025.

COLLABORATING FOR IMPACT

Working together to deliver meaningful projects

Arne ingkerreke apurtelhe-ileme (Gathering all the things together)

A collaboration with the Centre for Australian Languages and Linguistics (CALL) at Batchelor Institute of Indigenous Tertiary Education brought together the incredible life's work of Veronica Perrurle Dobson (AM). Veronica is an Eastern Arrernte elder, linguist, educator, author and ecologist, whose advocacy and work for language and culture is widely recognised and deeply respected. For over four decades, she has worked in many inter-related fields—language documentation and teaching, lexicography, interpreting, ethnobiology and land management. Her vast body of work is presented for the first time through *Arne ingkerreke apurtelhe-ileme* (Gathering all the things together).

The project website is a guide to Veronica's life, work and publications and contains links to archival collections of her work. This project provided a workflow for developing 'person-centred collections'. Many archival collections revolve around non-Indigenous collectors who have recorded and accumulated material over their lifetime. Person-centred collections reframe our focus and enhance our understanding of the collection items by showing them as the cohesive work of an Indigenous knowledge holder.

Visit the project website >



Image *Arne ingkerreke apurtelhe-ileme* project team: Angela Harrison, Camille Dobson, Veronica Dobson, Jennifer Green and Ben Foley. Photo by Charlee-Anne Ah Chee.

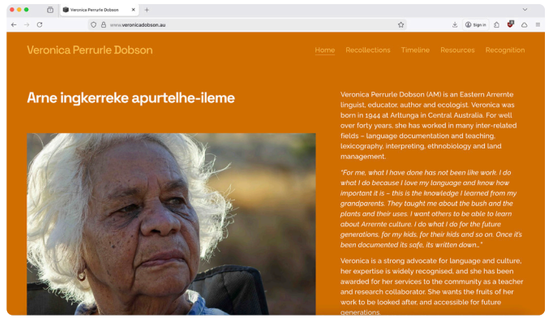


Image *Arne ingkerreke apurtelhe-ileme* website.



Image Gwen Warmbirrirr, Alex Hohoi and Jacinta Burukumalawuy at Djalkiri Keeping Place, Milingimbi.

First Languages Australia National Collections Strategy for First Languages

LDaCA and First Languages Australia (FLA) commenced work on the National Collections Strategy for First Languages. The project focuses on self-determination, community control and the right for First Language communities to access, use and store their language data. The project is grounded in principles of Indigenous Cultural and Intellectual Property rights and Indigenous Data Sovereignty and Governance.

FLA, with support from LDaCA, is leading the collections strategy, with the participation of community art, culture and language hubs, and state and national institutions.

In 2025, the LDaCA and FLA team undertook a listening and learning phase of the project. This involved travelling across Australia to learn about current contexts and approaches to community collection management. Alongside the primary work, the team also provided advice for individual collections and, in some situations, provided emergency data security services.

Outcomes of the strategy will include practical hands-on guides and recommendations for community organisations working with various aspects of collections, as well as a report to the Joint Council on Closing the Gap.

OUR WORK IN THE MEDIA

Bringing language research into the public conversation

Vulgar language

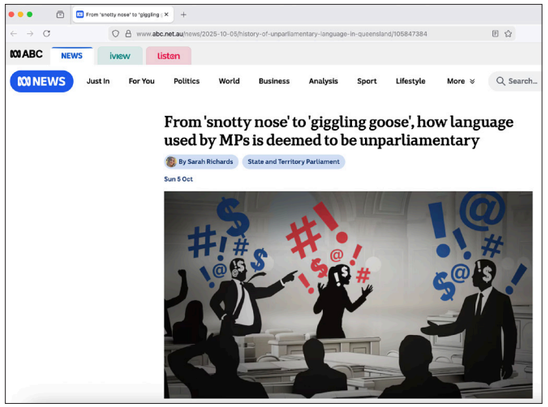
Research by CI Martin Schweinberger and colleague Kate Burridge (Monash University) on vulgar language in online spaces received national and international media coverage this year. Martin's and Kate's research uncovered who swears the most online by country—with surprising results! The journal article was published as part of a special issue of *Lingua* on bad language, edited by Martin, Kate and Paula Rautionaho (University of Eastern Finland). Martin was interviewed on ABC Drive Brisbane, ABC Drive Melbourne and 10News. The research made headlines in many media publications including The Conversation, CNN, The Guardian, Brisbane Times and on SBS.

Read the article >

Unparliamentary language in Queensland

In October, Sam Hames was interviewed by ABC News for an article about the history of unparliamentary language in Queensland. Sam explained how the language that is considered unparliamentary has evolved over time, reflecting broader changes in society.

Read the article >





SBS TV (Licensed by Copyright Agency)
News | AU | May 23 · 3:59 PM

【粗口研究】澳洲人講得少過英美？香港為最「文明」之一

3.76M Reach 13 Social Echo

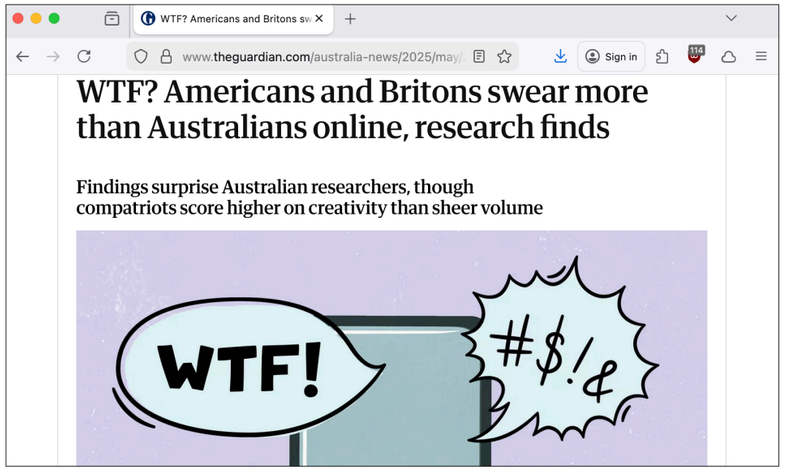


MSN.com · Joe Pinkstone
News | US | May 27 · 6:46 AM

Britain is world's second most potty-mouthed nation

110M Reach 0 Social Echo







Phys.org · University of Queensland
News | IM | May 23 · 10:17 PM

Global study finds Australians are third most prolific swearers

5.54M Reach 15 Social Echo

CONTRIBUTING TO SKILLS DEVELOPMENT

Supporting capability-building through workshops and training

ARDC Summer School & Indigenous Data Governance Masterclass

In February, we participated in the Australian Research Data Commons' (ARDC's) HASS and Indigenous Research Data Summer School, offering sessions in the FAIR and CARE principles, data management and computational analysis. Robert McLellan delivered an Indigenous Data Governance Masterclass with Levi Murray (previously Indigenous Data Network, University of Melbourne), sharing knowledge and tools for ethically managing Indigenous data with 160 attendees.

Read more >



Image Robert McLellan presenting at the presenting at the 2025 ARDC Indigenous Data Governance Masterclass. Photo by Marc Grimwade, supplied by the ARDC.

Graduate Digital Research Fellowships

LDaCA delivered its Graduate Digital Research Fellowship program for the third year in 2025, expanding its reach by moving fully online and welcoming fellows from across Australia. The cohort were supported by mentors from the LDaCA team and our partner organisation, QUT Digital Observatory. Our 2025 fellows had varied research topics, covering applied linguistics, heritage studies relating to language data and public policy in relation to alcohol.

Expressions of Interest for the 2026 Graduate Digital Research Fellowships are open now.

Learn more and submit an EOI for 2026 >

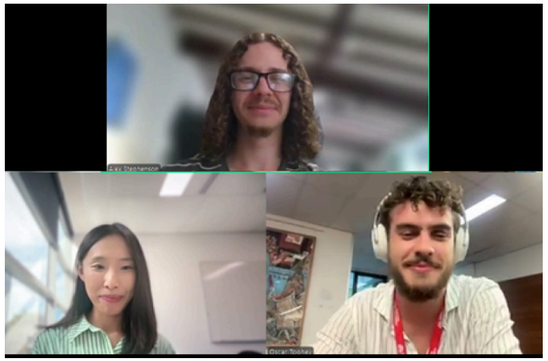


Image The 2025 GDRF Participants: Alex (top row), Elaine and Oscar (bottom row).



Image Ben Foley (LDaCA) presenting at the Darwin digital languages collections workshop.

Darwin Digital Languages Collections Workshop

From 24–28 March, the LDaCA team hosted a digital languages collections workshop in Garramilla/Darwin. This event brought together organisations and individuals from the Top End to exchange ideas and insights about Indigenous language collections and explore collaborative opportunities. During the week, we introduced participants to technologies, processes and practices to support their work with digital language collections and explored useful tools, such as the Crate-O platform, in hands-on sessions. Participants also engaged in a collaborative brainstorming session focused on metadata. The week concluded with rich conversations involving FLA, Libraries and Archives NT, Aboriginal Resource and Development Services (ARDS) and the Bininj Kunwok Regional Language & Culture Centre. These conversations explored sustainable strategies for managing language collections while prioritising community needs. The collaborative nature of this event laid the groundwork for a vibrant community of practice and highlighted the power of working together.

Read more >

SHARING KNOWLEDGE & STRENGTHENING NETWORKS

Sharing our work and learning from peers nationally and internationally

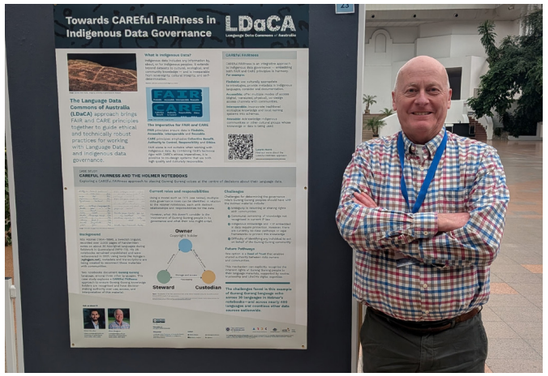
PULiiMA

We were a proud sponsor of PULiiMA 2025 (27–29 August), alongside the ARDC and welcomed attendees at our shared stall over the course of the conference. Our team also presented a themed session on ‘Language technologies and community-centred decision making’. The session aimed to demonstrate how modern tools and Indigenous-led approaches can strengthen connections between communities, language workers and institutions.



CLARIN Annual Conference

Robert McLellan and Simon Musgrave presented at our advisory partner CLARIN’s Annual Conference in Vienna, Austria in early October. Robert and Simon delivered a session about LDaCA and our work and a poster on Indigenous Data Governance. They also had the chance to engage with the international research community and meet with potential collaborators.









Images (left and above) LDaCA team members participating in PULiiMA and the CLARIN Annual Conference.

APPENDIX

LDaCA publications, workshops and presentations in 2025

Publications

Bednarek, Monika & Barbra Meek. (2025). "Are you Navajo or Inuit?" Identity, television dialogue and Indigenizing semiotics. *Journal of Linguistic Anthropology*, *35(1)*, e12449. **https://doi.org/10.1111/jola.12449**

Bednarek, Monika & Maite Taboada. (2025). Attitude in Reported and Non-reported News: A Critique of Sentiment Analysis in Corpus Pragmatics. *Corpus Pragmatics, 9(2)*, 111–133. **https://doi.org/10.1007/s41701-025-00185-6**

Hames, Sam, Michael Haugh & Martin Schweinberger. (2025). "How is that unparliamentary?": The metapragmatics of 'unparliamentary' language in the Australian Federal Parliament. *Lingua, 320*. **https://doi.org/10.1016/j.lingua.2025.103932**

Merx, Raphaël, Hanna Suominen, Lois Hong, Nick Thieberger, Trevor Cohn & Ekaterina Vylomova. (2025, May 24). TULUN: Transparent and Adaptable Low-resource Machine Translation. arXiv:2505.18683v1 [cs.CL]. **https://arxiv.org/abs/2505.18683**

Musgrave, Simon. (2025). Australian slang as a literary genre. *Australian Journal of Linguistics, 44(4)*, 411–430. **https://doi.org/10.1080/07268602.2024.2444219**

Nguyen, Li, Francesco Ricatti, Sacha Davis, Jaime W. Hunt, James A. Walker, John Hajek & Catherine E. Travis. (2025). Engaging with Oral Community Language Collections in Australia: Practices, Challenges and Ways Forward. *Journal of Intercultural Studies*, 1–18. **https://doi.org/10.1080/07256868.2025.2592305**

Passmore, Sam, Lila San Roque, Kirsty Gillespie, Saurabh Kumar Nath, Kira Davey, Keira Mullan, Tim Cawley, Jennifer Biggs, Rosey Billington, Bethwyn Evans, Nick Thieberger, Nicholas Evans & Danielle Barth. (2025). English-based acoustic models perform well in the forced alignment of two English-based Pacific Creoles. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 31172–31183, Vienna, Austria. Association for Computational Linguistics. **https://aclanthology.org/2025.acl-long.1505.pdf**

Puri, Anisa (2025). Oral History and Linguistics: Exploring Connections. *Studies in Oral History, 47*, 106–130. **https://oralhistoryaustralia.org.au/wp-content/uploads/2025/11/2025_OHAJ47_Article_Puri.pdf**

Reimann, Ritsaart, Kate E. Lynch, Stefan A. Gawronski, Jack Chan & Paul E. Griffiths. (2025). Classifying Genetic Essentialist Biases using Large Language Models. *Review of Philosophy and Psychology, 16(3)*, 1135–1165. **https://doi.org/10.1007/s13164-025-00770-3**

Schweinberger, Martin & Kate Burridge. (2025). Vulgarity in online discourse around the English-speaking world. *Lingua, 321*, 103946. **https://doi.org/10.1016/j.lingua.2025.103946**

Schweinberger, Martin & Michael Haugh. (2025). Reproducibility, replicability and robustness in corpus linguistics. *International Journal of Corpus Linguistics, 30(2)*, 119–129. **https://doi.org/10.1075/ijcl.25081.sch**

Schweinberger, Martin & Michael Haugh. (2025). Reproducibility and transparency in interpretive corpus pragmatics. *International Journal of Corpus Linguistics, 30(2)*, 234–260. **https://doi.org/10.1075/ijcl.23033.sch**

Thieberger, Nick. (2025). Avoiding semantic trespass and etyma-larceny. Collaborative lexicography for under-resourced languages, an example from Vanuatu. C. O'Shannessy, D. Angelo and J. Gray (eds), Projecting Voices: Studies in Language and Linguistics in Honour of Jane Simpson. Canberra: ANU Press, 593-611.

Thieberger, Nick. (2025). Records of Pacific Languages: Where Are They and Who Can See Them? Department of Pacific Affairs In Brief series. Canberra, ACT: Dept. of Pacific Affairs, Coral Bell School of Asia Pacific Affairs, The Australian National University **https://hdl.handle.net/1885/733771963**

Thieberger, Nick. (2025). The tip of the iceberg: Public good and the curation of humanities research records. *Arena Online*. **https://arena.org.au/public-good-curation-of-humanities-research/**

Workshops and conferences

JANUARY

McLellan, Robert & Levi Murray (IDN). (2025, February 3). *Indigenous Data Governance Masterclass* [Masterclass]. ARDC HASS and Indigenous RDC Summer School, Brisbane, QLD. **https://www.youtube.com/watch?v=fDUnUSsovBg**

FEBRUARY

Hames, Sam, Dan Angus (QUT) & Robert Fleet (QUT). (2025, February 4–5). *Obtaining and analysing HASS and Indigenous data* [Workshop]. ARDC HASS and Indigenous RDC Summer School, Brisbane, QLD.

Musgrave, Simon. (2025, February 4–5). *FAIR and CARE principles* [Workshop]. ARDC HASS and Indigenous RDC Summer School, Brisbane, QLD. **https://www.youtube.com/watch?v=UGGdLAeNWX8**

Musgrave, Simon, Ben Foley, Chenoa Pettrup, Rosanna Smith, Nooshin Laghai-Asiata & Alvin Sebastian. (2025, February 4–5). *Organising HASS and Indigenous data* [Workshop]. ARDC HASS and Indigenous RDC Summer School, Brisbane, QLD.

Foley, Ben. (2025, February 19). *Digital Technology for Language Growth* [Webinar]. First Languages Australia. **https://vimeo.com/1058458104**

Hames, Sam & Simon Musgrave. (2025, February 20). *Get started with Hansard for humanities, arts and social sciences research* [Workshop]. Melbourne Data Analytics Platform, Melbourne, VIC.

MARCH

Foley, Ben & Beau Williams (FLA). (2025, March 7). *Language collection management* [Conference presentation]. 9th International Conference on Language Documentation and Conservation, Honolulu, HI, USA.

Foley, Ben, Mosies Sacal Bonequi, Peter Sefton, Rosanna Smith & Chenoa Pettrup. (2025, March 24). *Darwin digital language collections workshop* [Workshop]. Darwin, NT.

APRIL

Foley, Ben. (2025, April 29). *Digital Technology for Language Growth* [Conference presentation]. Yunnigan Dhilla Yari Indigenous Languages Symposium, Brisbane, QLD.

MAY

Bird, Steven. (2025, May 29). *Language Technology and the Metacrisis* [Webinar]. The Language Data Commons of Australia.

JUNE

Bednarek, Monika & Maite Taboada. (2025, June 22–27). *Attitude (sentiment) and quotation in news discourse: A corpus pragmatic investigation* [Conference presentation]. 19th International Pragmatics Conference, Brisbane, QLD.

Hames, Sam & Eve Ansell (QCIF). (2025, June 22–27). *Interpretive text analytics and big data in pragmatics* [Conference presentation]. 19th International Pragmatics Conference, Brisbane, QLD.

Haugh, Michael, Sam Hames & Eve Ansell (QCIF). (2025, June 22–27). *What do people talk about in first conversations? Integrating computational and interpretive approaches to topic in initial interactions* [Conference presentation]. 19th International Pragmatics Conference, Brisbane, QLD.

Ober, Robyn, Tahlia Burchill-O'Brien, Diana Eades & Samantha Disbray. (2025, June 22–27). *Plenary Panel: The pragmatics of Indigenous languages* [LDaCA-supported event]. 19th International Pragmatics Conference, Brisbane, QLD.

JULY

Hames, Sam. (2025, July 1–3). *Introduction to Text Analysis with R* [Conference presentation]. Research Bazaar Queensland 2025, Brisbane, QLD.

Musgrave, Simon. (2025, July 1–3). *Exploring texts using concordances* [Conference presentation]. Research Bazaar Queensland 2025, Brisbane, QLD.

Arkadio, Abraham, Peter Akalkal, Nick Thieberger, Jamie Davidson & Jonathan Moodie. (2025, July 14–18). *Otuho Writers' Workshop* [LDaCA-supported event]. Melbourne, VIC.

Musgrave, Simon & Sam Hames. (2025, July 29). *Using large data sets for qualitative research: Examples from Australian Federal Hansard* [Webinar]. Universitas Udayana. **https://www.youtube.com/live/twCAdPazV2M**

AUGUST

Bednarek, Monika & Barbra Meek (University of Michigan). (2025, August 22). *Colour terms and Whiteness-marking in contemporary television series* [Seminar]. Research Seminar Series, Western Sydney University.

McLellan, Robert, Rose Barrowcliffe, Camille Dobson (Batchelor Institute) & Angela Harrison (Batchelor Institute). (2025, August 25–29). *Language technologies and community-centred decision making* [Conference presentation]. PULiiMA Indigenous Language and Technology Conference 2025, Darwin, NT.

Musgrave, Simon & River Smith. (2025, August 26). *Sustainable data for humanities researchers* [Workshop]. Melbourne Data Analytics Platform, Melbourne, VIC.

Sun, Chao & Alex Senhui Guo. (2025, August 26). *Explore, Analyse, Discover: Harness LDaCa Tools for Text Analysis* [Workshop]. Computational Social Science Lab, Sydney Corpus Lab & Sydney Informatics Hub, Sydney, NSW.

Erjavec, Tomaž & Matyáš Kopp. (2025, August 28). *The ParlaMint corpora of European parliamentary proceedings* [Webinar]. The Language Data Commons of Australia. **https://www.youtube.com/watch?v=SU92WFxmq-4**

SEPTEMBER

Pettrup, Chenoa, Michael Haugh, Simon Musgrave, Sam Hames & Robert McLellan. (2025, September 9). *Language Data Commons of Australia: Growing a mature research infrastructure ecosystem in HASS* [Presentation]. University of Queensland Research and Innovation Week, Brisbane, QLD, Australia.

Hames, Sam & Simon Musgrave. (2025, September 11). *Get Started with Federal Hansard for Humanities, Arts and Social Sciences Research* [Workshop]. Brisbane, QLD.

McLellan, Robert & Simon Musgrave. (2025, September 30 – October 2). *Towards CAREful FAIRness in Indigenous Data Governance* [Poster session]. CLARIN Annual Conference, Vienna, Austria. **https://bridges.monash.edu/articles/poster/Towards_CAREful_FAIRness_inTowards_CAREful_FAIRness_in_Indigenous_Data_Governance/30204508**

McLellan, Robert & Simon Musgrave. (2025, September 30 – October 2). *Language Data Commons of Australia: Towards a Nationally Distributed Research Infrastructure* [Conference presentation]. CLARIN Annual Conference, Vienna, Austria. **https://www.youtube.com/watch?v=1cARj6w3qZs**

OCTOBER

Musgrave, Simon. (2025, October 7). *Language Data Commons of Australia: Towards a Nationally Distributed Research Infrastructure* [Seminar]. The Endangered Languages Archive, Berlin, Germany.

Sun, Chao. (2025, October 10). *Language Data Commons of Australia (LDaCA) and Australian Text Analytics Platform (ATAP)* [Forum presentation]. ML4AU Tech Forum.

Barrowcliffe, Rose. (2025, October 13–16). *Reconnecting Indigenous Data to Country* [Keynote address]. SciDataCon 2025 (International Data Week), Brisbane, QLD.

Cooke, Steele (AARNet), Alex Ip (AARNet), Adam Bell (AARNet), Simon Musgrave & Rosanna Smith. (2025, October 13–16). *CAREful Linking of FAIR Language Data to Reproducible Jupyter Notebooks* [Poster session]. SciDataCon 2025 (International Data Week), Brisbane, QLD.

Haugh, Michael & Sam Hames. (2025, October 13–16). *The Language Data Commons of Australia: Supporting research for diverse communities* [Conference presentation]. SciDataCon 2025 (International Data Week), Brisbane, QLD.

Sacal Bonequi, Moises & Chenoa Pettrup. (2025, October 13–16). *The PILARS Approach: Designing Sustainable, Ethical and FAIR archival repositories* [Poster session]. SciDataCon 2025 (International Data Week), Brisbane, QLD. **http://doi.org/10.5281/zenodo.17240111**

Lynch, Michael, Eve Ansell (QCIF), Nichola Burton (ARDC), Ian McCrabb & Peter Sefton. (2025, October 20–24). *Sustainable Digital Humanities Collections* [Conference presentation]. eResearch Australasia 2025, Brisbane, QLD.

Sefton, Peter, Rosanna Smith, Simon Musgrave & Michael Lynch. (2025, October 20–24). *Low friction FAIR interoperability using RO-Crate metadata in text analytics pipelines* [Conference presentation]. eResearch Australasia 2025, Brisbane, QLD.

Sefton, Peter, Rosanna Smith, Nichola Burton (ARDC), Robert Fleet (QUT) & Michael Lynch. (2025, October 20–24). *Someone said to do RO-Crates? RO-Crate Birds of a Feather session meet colleagues and give feedback* [Conference presentation]. eResearch Australasia 2025, Brisbane, QLD.

NOVEMBER

Lee, Kelvin. (2025, November 4). *Artificial intelligence in corpus linguistic research* [Webinar]. Sydney Corpus Lab Dinner Talks.

Kemble, Melissa (USyd) & Monika Bednarek. (2025, November 7). *New tools for analysing Australian English: The case of emotion words and gender stereotyping* [Conference presentation]. 8th Forum on Englishes in Australia 2025, Melbourne, VIC.

Smith, Rosanna & Simon Musgrave. (2025, November 7). *Sharing the Australian Slang Survey Data collection* [Conference presentation]. 8th Forum on Englishes in Australia, Melbourne, VIC.

Schweinberger, Martin (2025, November 10). *Vulgarity in World Englishes* [Webinar]. University of South Australia. **https://www.youtube.com/watch?v=Y0EFJMZn4KM**

DECEMBER

Hames, Sam & Naomi Barnes (QUT). (2025, December 2–5). *Hacking the Hansard: A critically informed approach to digital heritage archival work* [Conference presentation]. Digital Humanities Australasia 2025, Canberra, ACT.

Thieberger, Nick. (2025, December 2–3). *The Written and the Recorded: AI in the workflow from analogue materials to textual corpora* [Conference presentation]. 2025 Canadian Australian Partnership for Open Scholarship (CAPOS) gathering, Canberra, ACT.

Hames, Sam & Simon Musgrave. (2025, December 2–5). *Get Started with Federal Hansard* [Workshop]. Digital Humanities Australasia 2025, Canberra, ACT.

Musgrave, Simon, Peter Sefton, Robert McLellan, Sam Hames, Ben Foley & Nick Thieberger. (2025, December 2–5). *Enriching the Ecosystem: LDaCA and DH in Australia* [Conference presentation]. Digital Humanities Australasia 2025, Canberra, ACT.

Sefton, Peter, Nick Thieberger & Michael Falk. (2025, December 2–5). *RO-Crate for a data commons* [Conference presentation]. Digital Humanities Australasia 2025, Canberra, ACT.

Barth, Wolfgang (2025, December 2–5). *Landing Pages for Linguistic Data Collections: Enhancing Access and Accountability in Variation Research* [Conference presentation]. Australian Linguistic Society Conference, Gold Coast, QLD.

Gnevsheva, Ksenia, Catherine Travis & Gerry Docherty. (2025, December 2–5). *What it means to be a local: The role of local attachment in vowel realisations in regional Australia* [Conference presentation]. Australian Linguistic Society Conference, Gold Coast, QLD.

Wood, Bronwyn, Catherine Travis & Gerard Docherty (2025, December 2–5). *Sounding Au[ʃ]tralian: Retraction of (str) across real and apparent time in spontaneous speech* [Conference presentation]. Australian Linguistic Society Conference, Gold Coast, QLD.

Acknowledgements

LDaCA wishes to acknowledge the collaborating institutions participating in the project

The Language Data Commons of Australia (LDaCA) (DOI: 10.3565/kq2v-9g52) is a co-investment partnership with the Australian Research Data Commons (ARDC) through the HASS and Indigenous Research Data Commons. The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).

LDaCA
Language Data Commons of Australia

ldaca.edu.au
ldaca@uq.edu.au
@LDaCA_Project
@ldaca