

# Data license A:

## Sydney Speaks: Polished transcriptions and audio (de-identified)

### 1. The parties:

- 1.1 **The Licensor:** Australian National University
- 1.2 **Data steward:** Catherine Travis, [catherine.travis@anu.edu.au](mailto:catherine.travis@anu.edu.au)  
School of Literature, Languages and Linguistics  
Australian National University
- 1.3 **The Licensee:** hereinafter referred to as 'You'.

2. The **Sydney Speaks** collection is managed by the Sydney Speaks project, a large-scale sociolinguistic project funded through the ARC Centre of Excellence for the Dynamics of English (2014-2022). The collection consists of contemporary audio recordings and transcriptions made between 2014 and 2023 and incorporates legacy data from two additional sources: a sociolinguistic project carried out in the late 1970s and an oral history collection created during the 1980s.

3. This licence refers to the de-identified polished, prosodic transcriptions and de-identified audio from the Sydney Speaks collection, hereinafter referred to as the 'Material'.

### 4. Data access conditions:

4.1 The Licensor grants You a limited, non-exclusive license to access and use the Material in accordance with the terms of this License Agreement. You may not sub-license, transfer or assign access rights to another party.

4.2 The Material may be used only for non-commercial purposes in accordance with the Approved Research Proposal. No additional communication, article or abstract based on the Material, nor any other use thereof, may be made without submitting a separate research proposal.

4.3 You are permitted to publish reports which result from your use of the Material as agreed upon in the Approved Research Proposal. A copy of the product of the research undertaken must be provided to the Licensor. All reports must include appropriate acknowledgment and references as per the instructions in Annex 1.

4.4 Examples from the collection must be reproduced verbatim as per the instructions provided in Annex 2.

4.5 You may not duplicate, circulate, share, or otherwise transmit the Material in any form (audio, electronic or print) to any third party.

4.6 You must exercise due professional care and judgement and only make use of the Material as permitted by this Licence and in accordance with the highest standards of

ethical and academic practice. The Material must not serve as the basis for personal judgements about the opinions, personality or language of the participants.

5. All intellectual property rights are reserved by the Owners and Licensees of those rights.
6. You must maintain adequate data storage measures to ensure the security of the Material.
7. After completion of the activities indicated on the Approved Research Proposal this License is terminated and all related Material, with the exception of strings for statistical analysis, must be discarded from personal computers and storage devices.
8. The Material is provided as is and the Licensor is not liable for the accuracy, content, legality, completeness, availability and reliability of the Material.
9. This Licence Agreement is governed by and to be construed in accordance with the laws of Australia and the Licensee will submit to the jurisdiction of Australian Courts.
10. The Licensor may terminate this License Agreement without notice.

Annex 1: Instructions for correct attribution of the Material.

The following attribution must be included in any publications, presentations and published reports:

The data on which this study is based were extracted from corpora provided by the Sydney Speaks project.

Publications, presentations and published reports should also refer the reader to Travis (2024) for more information about the participants, data collection protocol, and corpus management:

Travis, Catherine E. 2024. Sydney Speaks Corpus: An overview. *Australian Journal of Linguistics* 44(2-3):163-181. <https://doi.org/10.1080/07268602.2024.2386387>

Below are the recommended citation for each of the sub corpora included in the Sydney Speaks collection:

Citation for speakers from the legacy oral histories (collected 1987-1988): NSW Bicentennial Oral History Collection	Travis, Catherine E., James Grama, Simon Gonzalez, Benjamin Purser, & Cale Johnstone. 2023. Sydney Speaks Corpus. ARC Centre of Excellence for the Dynamics of Language, Australian National University. <a href="https://dx.doi.org/10.25911/m03c-yz22">https://dx.doi.org/10.25911/m03c-yz22</a>  NSW Bicentennial Oral History Project. 1987. <i>NSW Bicentennial oral history collection</i> : Council on the Ageing NSW Branch and NSW Oral History Association of Australia. Housed at the National Library of Australia
Citation for speakers from legacy linguistic recordings (Collected 1977-1981): Sydney Social Dialect Survey	Travis, Catherine E., James Grama, Simon Gonzalez, Benjamin Purser, & Cale Johnstone. 2023. Sydney Speaks Corpus. ARC Centre of Excellence for the Dynamics of Language, Australian National University. <a href="https://dx.doi.org/10.25911/m03c-yz22">https://dx.doi.org/10.25911/m03c-yz22</a>  1970s data (the “Sydney Social Dialect Survey”): Horvath, Barbara. 1985. <i>Variation in Australian English: The sociolects of Sydney</i> . Cambridge: Cambridge University Press.
Citation for speakers from the ANU Corpus of Sydney Speech	Travis, Catherine E., James Grama, Simon Gonzalez, Benjamin Purser, & Cale Johnstone. 2023. Sydney Speaks Corpus. ARC Centre of Excellence for the Dynamics of Language, Australian National University. <a href="https://dx.doi.org/10.25911/m03c-yz22">https://dx.doi.org/10.25911/m03c-yz22</a>

## Annex 2:

Examples from the Sydney Speaks collection included in presentations and publications and other materials must be reproduced verbatim, including punctuation to indicate prosodic information.

Each example must be attributed to its source by reference to the recording code attribution, and the beginning and end time stamps of excerpt in the format (hour:)minute:second.

e.g.     Dominic:         ... there is one,  
                              that .. goes for about .. a hundred metres,  
                              but that's it.                     (SSDS\_ITM\_106, 07:33-07:37)

The first example cited must include the following footnote:

Examples are produced verbatim from the transcripts. Within parentheses is the corpus, code and number, and beginning-ending time stamps of the excerpt given.