PERMISSION TO DISTRIBUTE TRANSCRIPT DATA
THROUGH TALKBANK
Carnegie Mellon University

Carnegie Mellon University is participating in a data-sharing program known as TalkBank that was established through federal and foundation grants to Brian MacWhinney at Carnegie Mellon. For the program to function properly, contributors will be asked to and will give permission for their work to be made available to other researchers. With a full understanding of the aforementioned, I hereby give permission to TalkBank to make and circulate electronic copies of the language transcripts and media that I describe below. These copies may be distributed to scholars and other responsible parties. I warrant that this use of the data is in accord with Human Subjects review procedures at my institution and that participants have given informed consent to have their data available to researchers. I also warrant that there is no copyright restriction over the transcripts and media being circulated. Any further restrictions that I wish to place on the use of these data are listed under (2) below. I do not hold Brian MacWhinney or Carnegie Mellon responsible for the enforcement of these further restrictions and indemnify and render harmless both Brian MacWhinney and Carnegie Mellon University against any actions at law or in equity or in similar courts of any jurisdictions arising from violations of these restrictions.

1. General description of the data set, and IRB Approval # (if available):

The Casillas HomeBank corpus includes language and other annotations for photo-linked audio recordings made with Tseltal families in Chiapas, Mexico and with Rossel families on Rossel Island, Milne Bay Province, Papua New Guinea. The Tseltal data were contributed in 2015 and the Rossel data is being prepared for contribution in late 2020. The audio recordings are originally in stereo .mp3 format, made with an Olympus WS-832 or Olympus WS-853 recorder. The photos (which are not yet posted as of November 2020) are in .jpg format, made with a Narrative Clip 1 camera. Associated annotations have been made in ELAN. More corpus details can be found here: https://docs.google.com/document/d/1KtmK2SiQfUig10AzUMp2BH3qNCZ1uVl5 eoSwlW833Is/edit. Data ownership is shared with the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, where the data are also archived. The data were collected under the approval of the Radboud University Social Sciences Ethics Committee (original approval: ECWS2013-1811-144 Holler/Levinson Fieldwork projects (L&C department, MPI); extension: ECWS2013-1811-144a1 Holler/Levinson Fieldwork projects (L&C department, MPI)). In Papua New Guinea, the data were also collected under collaboration with the National Research Institute of Papua New Guinea, who secured valid research visas for data collection, and the Milne Bay Province administrative office, who secured local researcher passes, both of whom maintain rights to copies of research output. If Marisa Casillas is not available to answer questions about this dataset, her co-authors (Stephen C. Levinson and Penelope Brown) should be consulted, along

with Elika Bergelson, who is highly familiar with the dataset and its intended use and share cases.

2. Restrictions to be placed on the use of the data:

These data should be kept in the Secure/Limited section of HomeBank (or any equivalent subset, should there be a future reorganization), meaning that anyone wanting access to the data must first (a) successfully get HomeBank membership and then additionally (b) submit a short form describing the intended use(s) of the dataset (currently: https://forms.gle/U1scft9PssJXmtFJ8) to be reviewed by Marisa Casillas or, in the event of her absence, by Elika Bergelson. In addition to these responses, applicants should answer the following questions and any other specific use questions generated by Casillas (or Bergelson, in her absence):
- What methods does your project have in place to make sure that all copies of the data are securely stored, tracked, and can be deleted at the end of the access period?
- If you need advice on appropriate references for this/these language(s) please contact Marisa Casillas, Stephen C. Levinson, and/or Penelope Brown.
- Would you be willing to, at time of publication or before, submit your annotated vocalization files back to the HomeBank corpus for re-use by yourself or others?
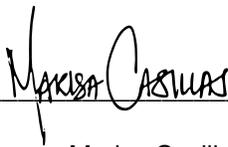
Applicants with specific and good-faith proposals (i.e., not harmful to the communities represented in any way, proximally in the data use or distally in anticipated use of the research outputs) who are also equipped with the infrastructure to support adequate data security should be given access approval for the specified project period by Casillas (or Bergelson, in her absence). Inappropriate data use, storage, or sharing and use, storage, or sharing inconsistent with what was approved gives Casillas (or Bergelson, in her absence) rights to revoke the user's permission to access the data.

Casillas corpus users must cite one or more of:
    Casillas, M., Brown, P., & Levinson, S. C. (2017). Casillas HomeBank Corpus. doi:10.21415/T51X12
    Casillas, M., Brown, P. & Levinson, S. C. (2020) Early language experience in a Tseltal Mayan village. Child Development, 91(5), pp. 1819–1835.
    Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Papuan community. Journal of Child Language 1-23. [please check at time of citation for additional issue and page number details]

Signed   _Marisa Casillas_   Date __25 November 2020__

Printed Name ___Marisa Casillas___   Institution __Max Planck Institute for Psycholinguistics__

Please send a signed copy to macw@cmu.edu or mail to: Brian MacWhinney, CMU-Psychology, 5000 Forbes Ave. Pittsburgh, PA, 15213, USA