

**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 WOCAE Corpus (Written and Oral Chinese Academic Emails) is a collection of emails written and read by Chinese students from Heilongjiang International University to their professor disagreeing about their results. The corpus is divided in two major groups: a written corpus and an oral corpus. The data for the written corpus was collected during the 2017/2018 academic year and was subdivided in two languages: Spanish and English. Each one is likewise subdivided in different groups depending on the academic year of the subjects. The Spanish-WOCAE Corpus groups 150 emails written by 2nd, 3rd and 4th year students, while the English-WOCAE Corpus groups 100 emails written by 2nd and 4th year students. The ages of the participants are between 18 and 27. The data for the oral corpus was collected in May 2018 and May 2019 and is divided in two languages: Spanish and Chinese. A total of 48 students of Spanish as a Foreign Language have participated in this corpus which was recorded in Harbin, China. The ages of the participants are all between 18 and 20, and they are all in the second year of their academic studies. The duration of each tape lasts between 44 seconds and 1:54 minutes.

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



Signed _____ Date 25/05/2020

Printed Name David Rodríguez Velasco Institution _____

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