**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. The *RusLan-M* corpus comprises longitudinal spontaneous speech samples of video recordings from two Russian-speaking monolingual children, a boy *Yasha* and a girl *Tosya*. Multimedia sources were anonymized by blurring all faces in the videos, muting or removing personal details (e.g., addresses and last names) in the transcripts, and replacing some videos with audio recordings to protect the participants’ privacy.

*Tosya* was recorded by her mother and caregiver from the age of 10 months, with the final recording made at 3;10. She lived in Russia and had an older sister. The total duration of Tosya’s recordings was approximately 29 h (21,421 child and 40,811 adult utterances), with approximately 36,275 and 126,944 tokens, respectively, where tokens represent all distinct, non-unique word forms.

*Yasha* was recorded by his father from the age of 1 year and 5 months, with the final recording made at 3;0. He also lived in Russia and had an older brother. The total duration of *the Yasha* recordings was approximately 12 h (13,965 child and 12,278 adult utterances), with approximately 27,592 and 41,141 tokens, respectively.

The transcripts of both child and child-directed speech were lemmatized and morphologically annotated using Mystem, an automatic tool for Russian morphological analysis (https://yandex.ru/dev/mystem). Trained research assistants then manually verified the annotations and resolved the cases of homonymy. Annotated tables containing nouns and verbs for both children are available on the OSF page (we will add them later).

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

There are no restrictions on the use of the transcripts.



Signed \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date \_02.03.2025\_\_\_\_\_\_\_\_\_\_\_

Printed Name \_\_Valeriia Lelik\_\_\_\_\_\_\_ Institution HSE University, Russia

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