

**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. For the program to function properly, many 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 Brian MacWhinney, Director of the Center of the Child Language Data Exchange System at Carnegie Mellon University, 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. 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.

The *Corpus PAROLE (PARallèle Oral en Langue Etrangère)* was compiled by members of the *Langages* research team at the Université de Savoie (Chambéry, France). It is composed of oral productions by 68 young adult learners of three foreign languages (English, French, and Italian), as well as a “benchmark” corpus of productions by 27 native speakers performing the same tasks. The corpus consists of audio files and the transcripts of each participant performing two short video summary tasks, and one short autobiographical narrative. In addition to these speaking tasks, all the non-native subjects completed a battery of tests and questionnaires, furnishing complementary data on their L2 knowledge and experience.

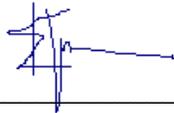
PAROLE was compiled in order to investigate the characteristics of learner language at different fluency levels. The particularity of the corpus is our attempt to include the temporal elements of spoken production in the main transcription line, along with more classic coding of errors and retracings.

Links to the video triggers used for the summary tasks, details concerning the participants, the corpus structure, the tests and questionnaires used for collecting additional subject data (and the data obtained from these instruments), as well as the specificities of our coding system are provided in the documents accompanying PAROLE in TalkBank. We welcome contact from any researchers interested in knowing more about our data or data collection procedure.

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

All participants in the *Corpus PAROLE* (who were remunerated at the minimum wage for the time devoted to the project) signed a consent form, authorizing the inclusion of the sound files and transcriptions of their spoken productions in a database accessible on the World Wide Web, to be used **for research purposes only**. The signed forms are on file at the Université de Savoie. All productions have been coded with an anonymous subject number; none of the subjects’ names (first, middle, or last) figure in any documents or transcriptions.

Signed: _____



Date: **15 July 2008**

Heather HILTON (project coordinator)
Laboratoire LLS, équipe *Langages*
UFR-LLSH, Université de Savoie
Chambéry, France
33/ 479 758 522 - hilton@univ-savoie.fr

Please mail this form to:
Brian MacWhinney, CMU-Psychology, 5000 Forbes Ave. Pittsburgh, PA, 15213, USA