The **MuSSeL Corpus**
A corpus of Mu(tilingual) S(poken) Se(cond) L(anguage) texts

The Second Language Teaching and Research Center at the University of Utah is developing a large-scale, multi-language learner corpus of second language speech.

**What is the MuSSeL Corpus?**
A learner corpus is a principled, computerized collection of learner language production. The texts in a learner corpus are carefully curated and can be digitally searched and/or analyzed in order to allow for large-scale investigations of their features, such as lexis, syntax, and language development. Such research contributes to our understanding of second language learning. In addition, research using learner corpora can inform language pedagogy, from materials development (e.g., definitions in learner dictionaries can be supplemented with information about frequent errors associated with that lexical item) to curriculum design and classroom methodologies (e.g., activities can be designed that anticipate and explicitly address frequent learner errors related to a particular lexical item or grammatical structure).

The MuSSeL Corpus project is focused on the development of a learner corpus containing spoken language produced by 1,800 learners of six foreign languages: Chinese, Russian, Portuguese, Spanish, French, and German. Speech samples are collected during testing using one of two instruments: Adult samples are collected using The American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview by computer (OPIc). Child samples are obtained from the ACTFL Assessment of Performance towards Proficiency in Languages (AAPPL). There are samples from each language in three contexts: longitudinal data collected from child L2 learners enrolled in Utah’s Dual Language Immersion programs in grades 3 and 5 (AAPPL samples), OPIc samples from adult L2 learners in classroom contexts, and OPIc samples from adult L2 learners who have acquired their L2 through immersion.

**Applications**
Researchers will be able to use the MuSSeL corpus to investigate a variety of questions related to second language learning and its associated mechanisms. These include the order of acquisition of linguistic features, the effects of first language transfer on L2 development, and the relative difficulty of particular linguistic structures in a given L2.

Teachers can use the corpus to inform their instruction in a variety of ways, including curriculum development and the development of lesson plans and activities. The corpus can be used to answer questions such as: What does learner speech look like at different proficiency levels? What types of difficulties do learners commonly encounter with a new vocabulary word or grammatical structure?

For additional information email: erin.schnur@utah.edu or visit https://l2treC.utah.edu/multi-Lingual_Speech_Corpus.php

We are grateful for seed funding from the VP for Research and the College of Humanities at the University of Utah, as well as funding from the Flagship Proficiency Initiative grant, and a Title VI Language Resource Center grant.