This brochure gives you an overview of the Language Corpora projects of COERLL, UT-Austin’s new Title VI language resource and research center. One of only 15 national foreign language resource centers, COERLL seeks to promote the culture of sharing at the heart of the Open Education movement. In more concrete terms, COERLL will produce and disseminate Open Educational Resources (OERs) for the Internet public (e.g., online language courses, reference grammars, assessment tools, corpora, etc.). The term OER refers to any educational material offered freely for anyone to use, typically involving some permission to re-mix, improve, and redistribute. And finally, in addition to embracing the tenets of Open Education, COERLL seeks to reframe foreign language education in terms of multilingualism. As such, all COERLL materials will strive to represent more accurately multilingual development and performance along multiple continua (dialectal, proficiency, etc.).
Language Corpora

Representing and Assessing the Bilingual Experience

Sociolinguistic corpora attempt to represent linguistic diversity by organizing authentic samples of a language as it is naturally spoken. As such, corpora have many potential uses for foreign language instruction. Unfortunately, corpora are typically built for researchers rather than teachers and students. The goal of this project is to produce a pedagogically useful corpus of Spanish and bilingual Spanish-English speech samples culled from interviews and naturalistic conversations among speakers of diverse personal profiles and regional origins throughout Texas. This open access website will include instructional materials for use in Spanish language classrooms and dual-language immersion programs, as well as Social Studies and Language Arts curricula.

Project director(s):
Dr. Almeida Jacqueline Toribio (Spanish & Portuguese)
Dr. Barbara Bullock (French & Italian)
Dr. David Birdsong (French & Italian)

* Project is in development

Spanish in the US

The United States ranks fifth with respect to number of Spanish speakers in the world. Half of the Spanish speakers in the US reside in Texas and California, where persons of Spanish-language heritage make up over one-third of the state populations. Texas leads in the percentage of residents who speak Spanish at home. But Texas Spanish speakers do not constitute a monolithic linguistic community; rather, they speak many different regional varieties and dialects from diverse sources: ancestral Spanish of the early settlers and national languages of immigrants from throughout the Spanish-speaking world. In addition, Spanish in Texas has been in sustained contact with English, leading to the conventionalized importation of English words, phrases, and grammatical patterns (Tex-Mex in the colloquial vernacular), but subtle effects of contact are in evidence even when speakers produce ‘monolingual’ Spanish utterances. Furthermore, Spanish language usage is not static; an individual may be dominant in one language but shift to another language later in life.

Sociolinguistic interviews

Speech samples for the corpus will be collected in different geographic locations around the state from speakers of diverse ages and social profiles engaged in guided but naturalistic interviews and conversations with Spanish-English bilingual interlocutors. Sessions will be videotaped in high-quality, uncompressed digital formats that will allow for the corpus to be exploited for a broad range of linguistic features. Speech samples will cover broad themes: from (im)migration stories to traditional folktales; from culinary recipes to instructions for speaking Spanglish. Participants will provide non-identifying information on their social characteristics and will be assessed for language proficiency, language dominance, language attitudes, and language usage. The videos will be edited into vignettes, coded for speaker information, and annotated with suppressible transcriptions.

Designed for students and teachers

The corpus will be used to develop materials that will enhance an appreciation of varieties of Texas Spanish and will permit parents, teachers, and practitioners to identify, assess, and understand normal variation. More important, the corpus will encourage teachers and students to view local varieties of Spanish in Texas as a resource to be exploited in the classroom and beyond, rather than as a deviation from some idealized norm. Lastly, documentation of Spanish language variation among speakers of diverse ages and social profiles will promote ideological shifts away from a deficit view of variation, thereby affording Spanish-speakers a more equitable participation in public discourse. The corpus will be supported by instructional materials that will facilitate incorporation into the classroom. Curricula will be designed for elementary, middle school, high school, and college students, and a manual will feature instructional tools, discussion questions, classroom and homework assignments, a bibliography, and links and references for other resources.

Assessing bilingualism

Many of the activities may be repeated on a regular basis, allowing for users to track trends and changes in patterns of Spanish language usage, attitudes, proficiency, etc. One of the research goals of the project is to document changes over time in relative language dominance among Spanish-speakers in Texas. For this purpose, the project will employ a brief, easy-to-use on-line questionnaire that will be made available to the public. This questionnaire and the information obtained from it will be of use to K-16 educators, administrators, language policy developers, and researchers who are interested in language proficiency in bilingual populations. Other questionnaires will address sociopsychological identity and Spanish language vitality in local communities, thus establishing measurable dimensions of language knowledge and use.