



NATIONAL ENDOWMENT FOR THE HUMANITIES

Parts of a Successful Application

The attached document contains the narrative portion of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and applicants are urged to prepare a proposal that reflects their unique project and aspirations.

Prospective applicants should consult the application guidelines at <https://www.neh.gov/program/dli-del-fellowships> for instructions.

Applicants are also strongly encouraged to consult with the NEH Research Division staff well before a grant deadline.

This attachment only contains the narrative, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

The application format might have been changed since this application was submitted. You must follow the guidelines in the currently posted Notice of Funding Opportunity (see above link).

Project Title: *Trilingual Diidxaza (iso:zai) - Spanish - English Dictionary*

Institution: University of Oregon

Project Director: Gabriela Pérez Báez

Grant Program: Dynamic Language Infrastructure - Documenting Endangered Languages (DLI-DEL) Fellowship

Diidxaza (aka: Isthmus Zapotec, zapoteco de la planicie costera; iso:zai, Glottocode: isth1244) is an endangered Indigenous Mesoamerican language of the Central Core Zapotec branch of the Otomanguean stock. The proposed project is to bring to completion a publishable manuscript of a 10,000-entry trilingual dictionary of Diidxaza with Spanish and English equivalents. This is the culmination of 20 years of field-based data collection and analysis and no further field research is needed. The performance period will focus on data synthesis and on writing Spanish and English equivalents for all 10,000 entries. The intended publication venue is *Dictionaria*, an online, open access and peer-reviewed dictionary journal.¹

The proposed dictionary contributes to the study of Mesoamerican languages. Diidxaza preserves the basic proto-Zapotec root structure based on consonant-vowel (CV) syllables, whereas most other Zapotec languages have lost vowels that precede or follow stressed syllables. Therefore, Diidxaza lexical data is essential to reconstruct ancestral forms of Zapotec languages (proto-languages). Further, Diidxaza data represent an important link between the earliest lexicographic documentation of Zapotec languages from the *Vocabulario en lengua çapoteca* from 1578 compiled by fray Juan de Córdova, and dozens of other contemporary Zapotec languages. This potential is tested in Pérez Báez and Cata (2021) which provides evidence, for instance, of continuity in the semantics of words and the native conceptualization of plants over 400 years. At the same time, it elucidates on the existence of three distinct prefixes used in the classification of botanical and other living referents. Thus, the publication of a trilingual Diidxaza-Spanish-English dictionary would have tremendous impact for comparative studies across Zapotec languages, across other superordinate genetic affiliations within the Zapotecan (Zapotec and Chatino) branch, and further across Otomanguean languages. English equivalents enable consultation by scholars around the world. Spanish equivalents further increase access by Latin American scholars especially in Mexico, and by heritage language community members. The unparalleled extensive scope of this dictionary and the online and open access availability of it will also benefit Diidxaza revitalization efforts.

Significance, impact, and endangerment. Diidxaza is spoken on the southern coast of the Isthmus of Tehuantepec in southern Mexico, and it is endangered. Marcial Cerqueda (2014) based on the Índice de Reemplazo Etnolingüístico (Ordorica 2009) shows that Diidxaza is undergoing *extinción lenta* ('slow extinction') to *extinción acelerada* ('accelerated extinction'). The Endangered Languages Project classifies Diidxaza as vulnerable with a speaker population of 80,000.² The Instituto Nacional de Lenguas Indígenas³ (INALI) classifies it at a *mediano riesgo de desaparición* ('medium/moderate risk of extinction').⁴ It reports, 122,697 from 2020 census data. However, there have been no studies to assess the level of proficiency among these speakers. Further, these numbers obscure the distribution of speakers across generations and the limited number of child speakers: of the 22 municipalities where Diidxaza is spoken, only in Santa María Xadani and San Blas Atempa are children growing up as speakers of the language (Marcial Cerqueda 2014). In the town of La Ventosa, where I have done most of my fieldwork, the youngest speakers are in their thirties. According to the 2000 census, in Tehuantepec which was once the capital of Zapotec communities in the Isthmus, only 4,434 speakers remain in a population of 53,226. Given the advanced language shift, there are local efforts underway to reclaim domains of language use for Diidxaza. Thus, the proposed dictionary will become a key resource for revitalization efforts.

Organization, concepts, and methods. The research for this dictionary began in the context of the Project for the Documentation of the Languages of Mesoamerica (PDLMA). Terence Kaufman began the research with Diidxaza speaker Javier López Cartas in 1995 and 1996 (as such, he will be recognized as co-compiler in the final publication). There was later a hiatus and research did not evolve consistently until I joined the project in 2003 and became the sole compiler. Within the PDLMA and through 2010, I worked with Rosalino Gallegos Luis and Rosaura López Cartas on data collection. I carried out additional

¹ <https://dictionaria.clld.org/>

² <https://www.endangeredlanguages.com/lang/1061>, last accessed April 26, 2023.

³ National Institute for Indigenous Languages in English, my translation

⁴ <https://atlas.inali.gob.mx/variantes/ficha/051362>, last accessed April 26, 2023.

field research under the auspices of two research projects. The Spatial Language and Cognition in Mesoamerica project involved experimental communicative tasks exploring body part terms, spatial language, metaphor and metonymy, and results were published in Pérez Báez (2011, 2018, 2019), Bohnemeyer et al (2015), Alsehri et al (2018), inter alia.⁵ I also directed an extensive ethnobotanical documentation project from 2012 to 2019 involving botanical data collection (cutting, photographing and processing plant samples) and linguistic and cultural data collection (recording of information sheets per plant, and audio recording, transcribing and translating plant descriptions). The results were published as the *La Ventosa Diidxaza Lexico-Botanical Dictionary* (Pérez Báez et al 2019) and other publications (Pérez Báez 2016, Pérez Báez and Cata 2021). In addition, I have carried out documentation and analysis of oral narratives (Pérez Báez et al 2015). All field research needed for the dictionary to be prepared for publication has been completed. I have also done data analysis and coding since 2010, yielding additional publications (Pérez Báez 2015, Pérez Báez and Kaufman 2016). The data has been managed in Toolbox and is currently being migrated to a SQLite relational database.

Diidxaza lexicographic works. Based on the above documentation efforts, the Diidxaza data for each entry is exceptionally thorough and of a scope that goes well beyond any other existing lexicographic works on Diidxaza. Picket et al (1956 and subsequent editions through 2013), and Jiménez Jiménez & Marcial Cerqueda (2000) have ~2000 entries each and only include the headword, a Spanish equivalent, and grammatical category (noun, verb, adjective, etc.). The latter dictionary builds upon the first and is to a great extent a reproduction of the first with the addition of sample sentences. Picket et al (2013) is available online.⁶ Jiménez Jiménez & Marcial Cerqueda (2000) is out of print. Jiménez Girón (1979) is a dictionary which is exceptional in that it is 1) handwritten, 2) the only one of all three compilations to orthographically represent tonal contrasts, and 3) includes verbal inflectional data in the habitual, progressive and future. However, it also has some 2000 entries only, and is out of print. There is one additional small pocket dictionary in Diidxaza and English (Britton 2003) with unfortunate mistranslations. Finally, a Diidxaza phrase compilation is available online at <http://www.biyubi.com>. A talking dictionary used to be available online at www.zapotecotecoc.org.mx but has been taken down.⁷

Scope of proposed dictionary. The proposed 10,000-entry dictionary far exceeds the scope of any past lexicographic works as it is about five times greater than any other Diidxaza dictionary published in the past. It is also at least twice the size of any of the lexical databases produced by the PDLMA for other Zapotec languages (except possibly the Chichicapan Zapotec database by Thomas Smith-Stark) and it will be the largest dictionary published by *Dictionaria* (the largest to date is the Tseltal-Spanish Multidialectal Dictionary with 8,109 entries). The proposed Diidxaza dictionary includes the following:

- **2,835** nouns
- **2,703** verbs, each with at least three inflectional forms in the habitual, potential and completive which have been used to code each verb for one of four verb classes recognized for Zapotec languages (cf. Pérez Báez & Kaufman. 2016); many entries have additional inflectional forms.
- **728** participles, each coded for any one of six types
- **237** adjectives
- **952** ethnobotanical entries incorporated from the *La Ventosa Diidxaza Lexico-Botanical Dictionary* (Pérez Báez et al 2019)
- Grammatical forms such as pronouns, demonstratives, adpositions, conjunctions, etc.
- Toponyms, denonyms, proper names, nicknames, collocations, and sound symbolic forms

All entries have grammatical category encoded, along with rigorous source attribution noting the name of the speaker who provided data and the corresponding year (speakers have agreed to have their names published), as well as references to any published source consulted to corroborate data. All headwords are

⁵ NSF Award No. BCS-0723694, PI J. Bohnemeyer.

⁶ https://www.sil.org/system/files/reapdata/47/57/69/47576984555300844250765337113555205633/zai_vocabulario_ed5.2.pdf, last accessed on May 17, 2023.

⁷ Last accessed on May 10, 2023.

written following the *Norma del sistema de escritura de la lengua zapoteca* (Cata et al 2016 ms). The dictionary also includes the technical PDLMA orthographic representation which is more faithful to the phonology of Diidxaza and rigorously documents prosodic stress patterns and root structure, and is compatible with some 30 other Mesoamerican language databases produced by the PDLMA. A sample entry with technical details is included in Attachment 5: Samples.

Competencies, skills, and access. Since 2003 I have carried out fieldwork-based documentation, analysis, and revitalization efforts for Diidxaza. I did so in the context of the PDLMA during the Summers of 2003, 2004, 2007, 2009 and 2010. In addition, I have carried out independent field research in the community of La Ventosa numerous times since 2006. Of particular relevance is the highly collaborative process I led in La Ventosa for the compilation of the *La Ventosa Diidxaza Lexico-Botanical Dictionary* (Pérez Báez et al 2019), a specialized, encyclopedic and trilingual Diidxaza-Spanish-English ethnobotanical dictionary (Pérez Báez et al 2019).⁸ This project, which ran from 2013 to 2019, involved 12 collaborators of which 7 are members of the Diidxaza language community; these members carried out critical active roles in project design and decision-making, data gathering and curation, knowledge transmission, and writing. The data collection involved 21 community members. This extensive fieldwork has led to numerous descriptive and theoretical works on Diidxaza on topics such as verbal morphology, semantic typology, and spatial language and cognition. These have been mentioned in this narrative and are listed in Attachment 4: Resume which include publications in high profile peer-reviewed publications such as *Cognitive Science*, many co-authored with members of the Diidxaza community, as well as national and international conference presentations, plenary talks, public presentations, many of which have also been collaborative. These works are considered key reference materials for the study of Diidxaza.

Final product and dissemination. The dissemination plans for the 10,000-entry trilingual Diidxaza-Spanish-English dictionary are informed by the experience publishing the *La Ventosa Diidxazá Lexico-Botanical Dictionary* (Pérez Báez et al 2019) in *Dictionaria*, a peer-reviewed, electronic, open access dictionary journal of the Max Planck Institute for Evolutionary Anthropology (MPI). I have been in conversation with the editors (b) (4), who are greatly enthusiastic at the prospect of publishing the proposed dictionary as it would be their largest dictionary publication to date. Dictionaries in *Dictionaria* are published as searchable, updatable databases. The institutional infrastructure of the MPI protects the data and ensures longterm access to it.

Online access is important to the language community whose internet connectivity and smartphone use have expanded considerably in the last decade. Further, the open access nature of *Dictionaria* allows for physical publication of the dictionary, in addition to the electronic version online, without special copyright permissions. Therefore, as part of the proposed project, I will produce a LaTeX output of the dictionary in PDF for ease of download and for print publication.

Dictionaria ensures peer-review of its publications. Given the extensive nature of the 10,000 entry dictionary, in consultation with *Dictionaria*, we devised a two-phase peer-review process. The first was a submission in June 2023 of 1% of the entries, with representation of all types of entries, to ensure they fit the *Dictionaria* data structure. The second phase will be in June 2024 with a 10-15% sample of the entries for external peer-review. (cf. Attachment 2: Work Plan for further details). I will be working on this peer-review dataset during Spring 2024 under a 10-week fellowship sponsored by the Oregon Humanities Center at the University of Oregon. Finally, (b) (4)

has expressed interest in publishing the proposed dictionary alongside the digital version of the *Vocabulario en lengua çapoteca* (Córdova 1987).⁹ If this manifests, it will increase the impact of the proposed dictionary on the advancement of historical reconstruction of Otomanguean languages and for language typology.

⁸ <https://dictionaria.eild.org/contributions/diidxaza>

⁹ <https://www.iifilologicas.unam.mx/cordova/index.php>