

University of New Mexico

UNM Digital Repository

Shared Knowledge Conference

Nov 8th, 11:00 AM - 1:00 PM

Creating a corpus of Navajo Historical Narratives - Prospects and Challenges

Lukas Denk

University of New Mexico

Follow this and additional works at: <https://digitalrepository.unm.edu/skc>

Denk, Lukas. "Creating a corpus of Navajo Historical Narratives - Prospects and Challenges." (2021).
<https://digitalrepository.unm.edu/skc/2021/Posters/23>

This Event is brought to you for free and open access by UNM Digital Repository. It has been accepted for inclusion in Shared Knowledge Conference by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Creating a corpus of Navajo Historical Narratives – Prospects and Challenges
Lukas Denk, PhD candidate in Linguistics, University of New Mexico
In collaboration with the Navajo Program at UNM and the Center for Regional Studies

Linguistic research and language instruction have benefitted from linguistic corpora (digital databases containing numerous texts written in one language). For low-resource languages where written records are scarce, the creation of corpora can be a challenge, but has the benefit of supporting conservation and revitalization. My project consists in the creation of a corpus of Diné Bizaad (Navajo), based on existing publicly available narratives written in the 1950's.^{1,2}

I present examples from the corpus as well as mention challenges faced in the annotation of the first 5000 words. One problem that pertains to Native American Languages in general is that only few speakers are trained in linguistic analysis. Learning how to disentangle the meaning in words is necessary for the creation of a grammatically annotated corpus, and it requires long training, offered only at certain institutions. Another challenge results from the particular structure of the language: even for linguists, the annotation of Navajo words is not straightforward, since grammatical rules are word-specific. Unlike in English or Spanish, word components in Navajo comprise both lexical and grammatical information, which means that every word has a different grammatical pattern. Resources like the Navajo Dictionary³ and the Analytical Lexicon⁴ provide support in resolving ambiguities in annotation. On the other hand, students and instructors of the Navajo Language Program at UNM have sufficient expertise to collaborate in this project. Using the free software FieldWorks Language Explorer and the depository Language Depot, the corpus can be accessed, downloaded and expanded. The goal of building this corpus is to facilitate empirical research on Navajo and to improve education by providing enough data that can be used to develop exercises and assignments.

¹ Young, R. W., & Morgan, W. (Eds.). (1952). *The Trouble at Round Rock*. US Department of the Interior, Bureau of Indian Affairs, Division of Education.

² Young, R. W., & Morgan, W. (Eds.). (1954). *Navajo historical selections* (No. 3). Phoenix Indian School Print Shop.

³ Young, R. and W. Morgan (1987). *The Navajo Language*. Albuquerque, University of New Mexico Press

⁴ Young, R., W. Morgan, S Midgette (1992). *An Analytic Lexicon of Navajo*. Albuquerque, University of New Mexico Press.