CREATING A CORPUS OF NAVAJO NARRATIVES
Lukas Denk (Department of Linguistics) at the UNM Shared Languages Conference, November 8th 2021

PROJECT
• Digital morpheme- and word-level annotation (1) of Navajo narratives
• Current: ~ 5,000 words, 4 narratives.
• Goal: 10,000 words by Spring 2022
• Includes translation, lexicon and a grammatical sketch

NARRATIVES
• The historical narratives are told by Navajo speakers and transcribed by Young & Morgan in the 1950s. They depict stories of quotidian life, real events, and stories. In contrast to the federal Indian Life Readers, these narratives depict a familiar, unmasked and vivid Navajo reality, a literary novelty for its time.
• Young & Morgan 1952: “The Trouble at Round Rock” (2)
• Young & Morgan 1954: “Navajo historical selections” (3)

SUPPORT
• Annotation: FieldWorks Language Explorer (4)
• Sources Dictionary (CD version) (5); word-form based search, contains paradigms and root forms, Analytical Lexicon (6, 7): root-based search, contains verb roots with information about their possible shape for tense, mood and aspect, their conjugation patterns and other prefixes
• Dr. Melvatha Chee as a consultant

CHALLENGES
• Digitalization of Navajo material in its infancy. Lack of uniformity in annotation/glossing, and in breaking down words into smaller entries.
• Recent history of writing only dates back to 1849, and narratives only appear to be collected in the 20th century. This means that there is not a huge collection of texts, and less possibilities to compare change within texts.
• Need to train Native speakers for annotation and contribute to a growing corpus
• In Navajo, word-specific grammatical rules make it relatively difficult to disentangle the grammatical and lexical information in words. There are numerous conjugation patterns and each verb selects its own prefixes; annotation might entail lengthy investigation of possible meanings (4).

PROSPECTS
• Narratives and software are open-source/free of copyright.
• Better access to Navajo literature and language
• More empirical research on the Navajo language, such as: “How often does the morpheme '/át/ vs. '/it/ occur in the language? When do speakers use jiní?”
• Input for further computational purposes (such as semantic annotation programs)
• Enriching the teaching landscape by providing various examples of spoken language. Instructors can easily obtain glossed examples for exercises and find words in their context
• Morpheme- and word-glossing improves the understanding of the relation between literal and figurative meanings, substituting searches in lexicon and grammar.
• Free access through Language Depot (https://public.languagedepot.org/nav-na-flex). Corpus can be modified and updated by instructors, researchers and students, making use of the growing lexical database.