Blackfoot Words database: Reflecting on our methods and projects

Background Blackfoot Words is a relational database of words and phrases in Blackfoot (Weber et al. 2023). The published version (v1.1; 2022-07-11) includes 4,553 inflected word tokens across 9 sources. To date, we have digitized words from over 50 sources, representing all four major dialects, and spanning the years 1743–2017. In this talk we reflect on our methods of analysis and provide suggestions for other researchers creating a database. We also describe some of our current projects.

Methods   The words in the database differ in orthography and morphological complexity. In order to provide links between parts of words, we analyzed inflected words into stems and morphemes and linked instances of the "same" stem or morpheme to a common lemma in a standardized orthography.
        There are two bottlenecks in our workflow. First, because words are not linked to lemmas, we require all words to be stemmed before they are entered into the database. Second, the morphemic analysis into stems and morphemes has fallen solely to the project director. While the first issue could be fixed by allowing words to be linked to lemmas, the second issue requires either a dedicated program or computational power to assist the research assistants.

Current projects   Our main project investigates morphophonological alternations within the stem in Blackfoot (e.g., between initials, medials, and finals). This supports a typological study of variation in prosodic structure across Algonquian languages. We also report on two collaborative projects. One investigate variation within Blackfoot, while the second creates computational tools like morphological parsers of Algonquian languages.