

Blackfoot Words: Introducing a database of Blackfoot lexical forms

Natalie Weber, Evan Hochstein, Pinyu Hwang, Nico Kidd, Diana Kulmizev, Hannah Morrison

Yale University

Overview

This presentation describes the structure and creation of Blackfoot Words, a new relational database of words and phrases in Blackfoot (ISO 639-3: bla). To date, we have digitized 62,693 individual lexical forms from 26 sources, representing all four major dialects, and spanning the years 1743–2017. This database was created by the Blackfoot Lab at Yale to provide digital access to the lexical forms in these resources, which are often difficult to locate or obtain, and to create links between words which contain the same lemma. We expect that the database will support projects by the language community and other researchers.

Challenges

The sources differ in their (1) orthography, (2) maximal stems (e.g. words minus inflectional morphology), and (3) analyses of stem components, if given. This non-uniformity of the data obscures the connections between related forms. To address these challenges, we analyzed inflected words into stems and morphemes ourselves and linked instances of the “same” stem or morpheme to a common lemma.

Database structure

The database comprises five tables: Sources, Words, Stems, Morphemes, and Lemmas. The Source table contains bibliographic information and commentary on the sources. The Words table contains inflected words in the source orthography. Each word is broken down into stems and morphemes which are entered into the Stems and Morphemes tables in the source orthography. The Lemmas table contains abstract versions of each stem or morpheme in a standardized orthography. Instances of the same stem or morpheme are linked to a common lemma.

Intended projects and collaborations

The database structure can accommodate lexical forms from new types of sources (e.g. audio recordings) as well as additional fields to support future projects. We end by elaborating some of our research goals and collaborations, including studies of dialectal variation, inflectional morphology and parsers, and historical change.