Blackfoot Words

Introducing a database of Blackfoot lexical forms

Natalie Weber, Evan Hochstein, Pinyu Hwang, Nico Kidd, Diana Kulmizev, and Hannah Morrison (Yale University)

October 15, 2021

53rd Algonquian Conference

Overview of Blackfoot Words

- relational database of words and phrases, and their subparts
- 62,693 lexical forms have been digitized to date
- from 26 sources
  - all four major dialects
  - timespan: 1743–2017
- Version 1 includes 7 of 26 sources

(for Blackfoot: Frantz 2017; Frantz & Russell 2017)

Motivation

1. Research needs
   - Blackfoot is unusual w.r.t. Algonquian language family
     i. Contains *archaisms* and *innovations* (Berman 2006; Goddard 1994, 2018)
     ii. Many unknowns, incl. variation, morphological composition of words
   - Each source contains only a subsets of the total number of lexical items

2. Teaching needs
   - Aging speaker population
     i. ~5,000 reasonably fluent speakers (= 15% population) (Genee & Junker 2018)
     ii. Language status: "shifting" (Canada), "moribund" (U.S.) (Ethnologue; Eberhard et al. 2021)
   - Active language maintenance (e.g. Piegan Institute in MT, language classes, etc.)

Purpose

To create an accessible, organized resource which can potentially support research projects as well as community-based language maintenance projects.

Two problems
- older sources are difficult to access or discover
- variation across sources means it is unclear when two words are the "same"

Specific aims:
1. to provide access to legacy sources by *digitizing* the inflected forms within
2. to provide links between instances of the "same" lexical form

Overview of talk

1. Data

2. Challenges & Decisions

3. Database Structure

4. Accessing the Database

5. Methods

6. Current & Future Projects

Data

Data: Overview

- Data includes variation in:
  - Dialect
  - Source (type)
  - Grammar (amount of complexity)
  - Orthographies

Data: Dialects

- Differences in:
  - pronunciation
  - morphology
  - lexical items (words)

- Not all sources specify dialect or speakers

(Map by Kevin McManigal)

(Bliss & Ritter 2009; Naoki 2014; Frantz 8 Russell 2017; Taylor 1969)

8

Data: Sources

- most are published, all are typeset
- almost 300-year span
  - oldest: Isham's (1743) wordlist of 10 numerals
  - most recent: *Blackfoot Grammar* (Frantz 2017)
- wide range of authors
  - missionary linguists
  - explorers for fur trading companies
  - ethnologists/anthropologists
  - philologists/linguists
  - amateur enthusiasts
- most sources with 1,000+ records document Aamsskáápipikani (So. Piegan)
- many different types (grammars, dictionaries, wordlists)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes
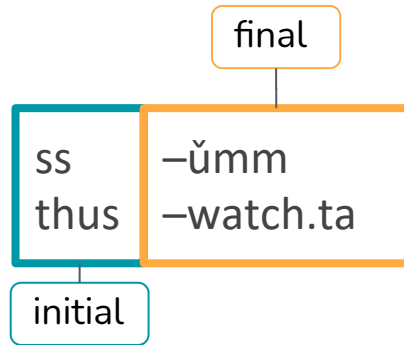
nitŭs'sŭmmosi

'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes
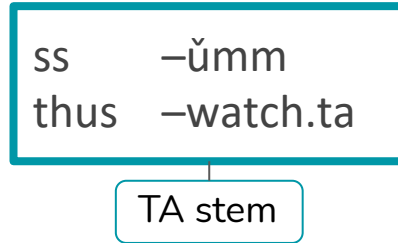
nitŭs'sŭmmosi

'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
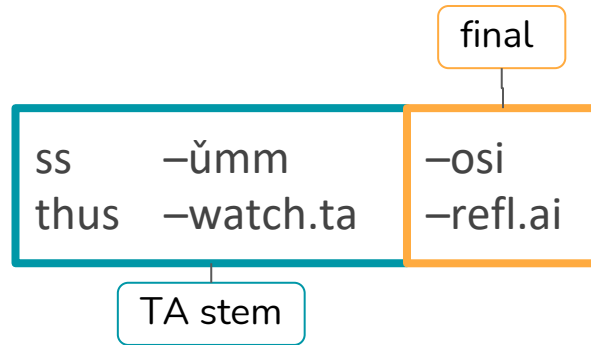- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

ss	–ŭmm
thus	–watch.ta

TA stem

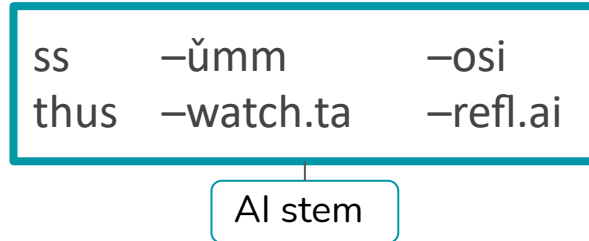'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

		final
ss	–ŭmm	–osi
thus	–watch.ta	–refl.ai

TA stem

'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
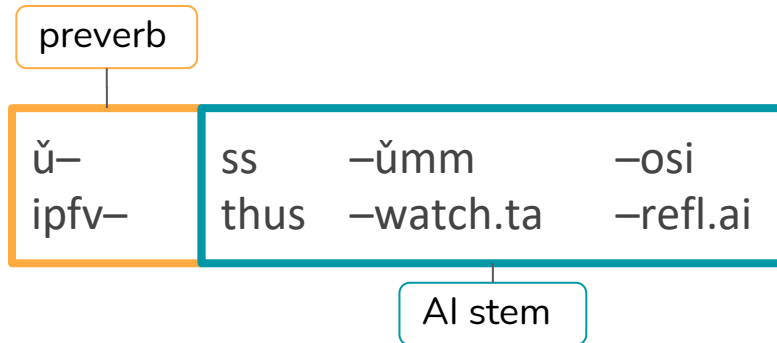- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

ss	–ŭmm	–osi
thus	–watch.ta	–refl.ai

AI stem

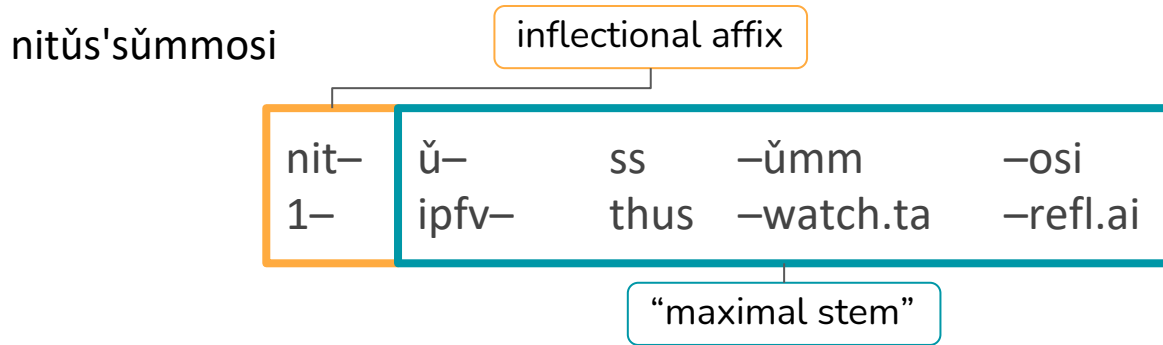'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

'I see myself.' (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

ŭ–	ss	–ŭmm	–osi
ipfv–	thus	–watch.ta	–refl.ai

"maximal stem"

'I see myself.'   (Tims 1889)

Data: Grammar

- Polysynthetic language with "clausal" words (Bliss 2013; Weber 2020, 2021)
- Stems can be recursive (Bloomfield 1946)
- Stems combine with many preverbs
- "Maximal stem" (our term): all material except for inflectional affixes

nitŭs'sŭmmosi

inflectional affix

| nit–
1– | ŭ–
ipfv– | ss
thus | –ŭmm
–watch.ta | –osi
–refl.ai |

"maximal stem"

'I see myself.'

(Tims 1889)

Data: Orthography

s**aah**kómaapi	'boy'	Frantz & Russell (2017: 232)
S**ah** komape	'Boy'	Catlin (1842)
s**a**coomahpa	'boy'	Latham (1846)
s**a**-ko'-ma-pi	'boy'	Hayden (1863)
s**o**k ko mA pe	'a boy'	Lanning (1882)
S**ar**komâpi	'Boy'	Lacombe (1886)
s**ŏ**k-u-ma-pi	'boy'	Tims (1889)
s**aχ**kúmapi	'boy'	Uhlenbeck (1938)

Challenges & Decisions

Challenges & Decisions

- **Tokenization:** breaking data into meaningful linguistic units

- **Phonemicization:** standardizing orthographic variants

- **Lemmatization:** determining abstract forms of linguistic units

Challenges & Decisions: Tokenization

- Tokenize at all linguistically relevant levels (words, stems, morphemes)

Inflected word:

nitǔs'sǔmmosi 'I see myself'

(Tims 1889)

(inflection: nit- '1')

Tokenization

Max stem:	[ǔ-	[s'sǔmmosi	]]
Category:	[preverb	[VAI	]]
Precedence:	1	2	

Stem:		[[s'sǔmm	]	-osi	]
Category:		[[VTA	]	fai	]
Precedence:		1		2	

Min stem:		[s's-	ǔmm	]
Category:		[init	fta	]
Precedence:		1	2	

Challenges & Decisions: Phonemicization

- (Extreme) variation in data, within and between sources

s**o**k ko mA pe   'a boy'   Lanning (1882)

S**ar**komâpi   'Boy'   Lacombe (1886)

s**ǒ**k-u-ma-pi   'boy'   Tims (1889)

s**aχ**kúmapi   'boy'   Uhlenbeck (1938)

Challenges & Decisions: Lemmatization

- Phonemicized at an abstract lemma level
- Lemmas link all instances of the same stem or morpheme

s**o**k ko mA pe	'a boy'	Lanning (1882)
S**a**rkomâpi	'Boy'	Lacombe (1886)
s**ǒ**k-u-ma-pi	'boy'	Tims (1889)
s**a**χkúmapi	'boy'	Uhlenbeck (1938)

271
saahkómaapi
NA
'boy'
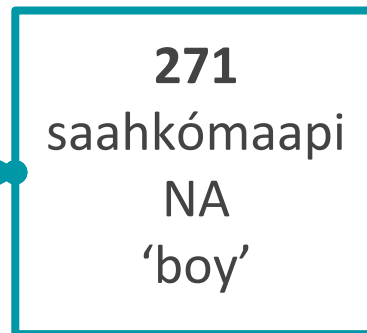
Challenges & Decisions: Lemmatization

Inflected word: **nitŭs'sŭmmosi 'I see myself'** (Tims 1889)

(inflection: nit- '1')

Tokenization

Max stem:	[ŭ-	[s'sŭmmosi		]]
Category:	[preverb	[VAI		]]
Precedence:	1	2		
Stem:		[[s'sŭmm	]	-osi]
Category:		[[VTA	]	fai]
Precedence:		1		2
Min stem:		[s's-	ŭmm]	
Category:		[init	fta]	
Precedence:		1	2	

Lemmatization

Stems

aissammohsi	VAI	's.o. sees themself'
ssammohsi	VAI	's.o. sees themself'
ssamm	VTA	's.o. sees s.o.'

Morphemes

a-	pv	'ipfv'
-ohsi	VAI	fai, 'refl'
ss-	init	'thus'
-amm	fta	'watch'

Database Structure

Database Structure

- Structure emerged as a response to challenges seen in previous slides
- Structure captures
  a. hierarchical structure of stems
  b. abstract relationships between stems and their counterparts
- Separates the original source data from the lab's analysis

Words

Word_ID
OriginalWord
WordTranslation
Dialect
Page
Source_ID
OriginalPhrase
OriginalPhraseTranslation

Morphemes

Morpheme_ID
Stem_ID
LabMorpheme
LabMorphemeCategory
Lemma_ID
Precedence

Stems

Stem_ID
ContainedIn
LabStem
LabStemCategory
Word_ID
Lemma_ID
Precedence

Sources

Source_ID

Lemmas

Lemma_ID
LabLemma
LabLemmaCategory
LabLemmaTranslation

27

Tables

1. Sources
- Bibliographic information as well as notes on dialect, orthography, provenance

2. Words
- Unit of record is lexical phonological forms
- Phrases are "chunked" into word tokens

3. Stems & 4. Morphemes
- Where the lab imposed tokenization ("chunking") on the forms in the words table
- Uses the source orthography

5. Lemmas
- Abstraction over multiple tokens of the same stem/morpheme across sources and orthographies
- Uses a standardised orthography (Frantz 1978, 2017)

Accessing the Database

Accessing the Database

All software used to create the database is free and open-source.

The full database is open-access and downloadable.

- Project website: https://www.blackfootwords.com/
- Interactive spreadsheet API (NocoDB): https://www.blackfootwords.com/view/
- MySql database and API hosted on Yale Spinup
- Downloadable (mysqldump) via Zenodo (coming soon!)
- Paper describing the database structure (ask us!)

Blackfoot Words

A database of lexical forms

About   How-to   **View**   Sources   Credits
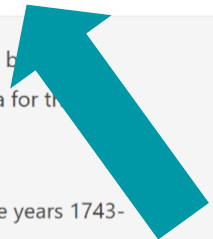
Blackfoot Words is a database of lexical forms in Blackfoot (Algonquian). By "lexical forms" we mean inflected words, stems, and morphemes. These have b collected and digitized from many different written sources. We created the database and this website to provide access to a large amount of lexical data for t Blackfoot communities and for language researchers.

Version 1 of the database includes lexical forms from legacy language documentation materials, including grammars, dictionaries, and wordlists, from the years 1743-2017.

Check out our How-to section to learn how to log in and View the database. Developers who are familiar with MySQL can Download the full database from Zenodo.

We have bibliographic information for all of the Sources in the database, as well as links to download pdfs of sources in the public domain.

The database was created by the Blackfoot Lab at Yale with the support of many others. The project is maintained by Natalie Weber. The language and words belong to the Blackfoot Nations.

Land acknowledgement

Window Snip
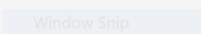
The database is hosted on a Yale-affiliated server. Yale University acknowledges that indigenous peoples and nations, including Mohegan, Mashantucket Pequot, Eastern Pequot, Schaghticoke, Golden Hill Paugussett, Niantic, and the Quinnipiac and other Algonquian speaking peoples, have stewarded through generations the lands and waterways of what is now the state of Connecticut. We honor and respect the enduring relationship that exists between these peoples and nations and this land.

31

NocoDB (*cont.*)

NocoDB (*cont.*)

NocoDB (*cont.*)

Single-record view
(from the **Words** table)

Words (iiksíkimmapiipitsi?wa)

A OriginalWord

iiksíkimmapiipitsi?wa

A OriginalTranslation

he is always very kind, he is a very kind person

A LabWordCategory

V

⊞ Words => Stems

🔗 Link to 'Stems'

iiksíkimmapiipitsi (Primary Key : stem-015)

íkimmapii (Primary Key : stem-016)

‹ **1** ›

NocoDB (*cont.*)

NocoDB (*cont.*)

(The **Lemmas** table)

Lemmas (poon-)

A LabLemma

poon-

A LabLemmaCategory

init

A LabLemmaTranslation

break, smash

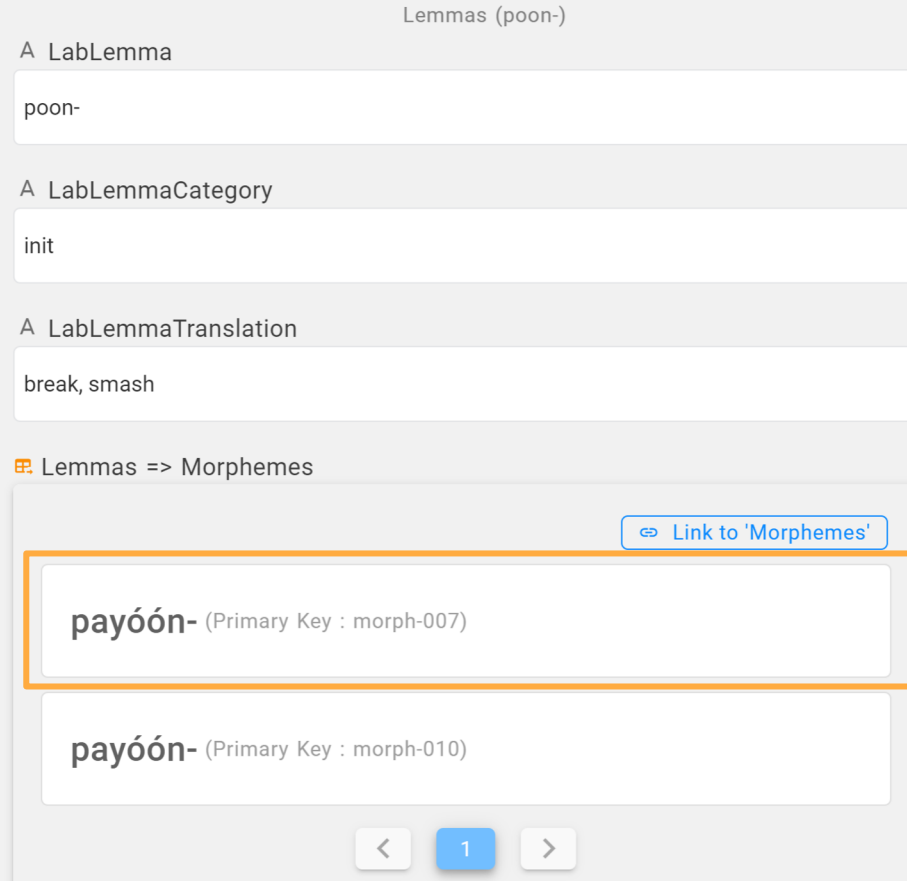⊞ Lemmas => Morphemes

🔗 Link to 'Morphemes'

payóón- (Primary Key : morph-007)

payóón- (Primary Key : morph-010)

⟨ 1 ⟩

NocoDB (*cont.*)

NocoDB (*cont.*)

Sorting (below) and **column/field selection** (right)

NocoDB (*cont.*)

Filtering (above)

Methods

Methods

Four phases:

1. Discovery   (tool: HathiTrust)
2. Digitization   (tool: Google Drive)
3. Analysis of Maximal Stems, Lemmas
4. Analysis of Non-Maximal Stems, Morphemes, Lemmas

- Always double-checked by a student, then by lab director
- Different sources progress at different rates
- Batch import into database after stages 3 and 4

Phase 2: Digitization

Source	Page	Speaker	Dialect	Word_ID	OriginalWord	WordTranslation	Original Category	OriginalUR	LabWord Category
AT1969	154	Dave Melting Tallow (North Piegan)	Aamsskáápipikani (Southern Piegan)	word:AT1969:1511	koxkátsinnawanists	our legs			N
AT1969	144	Margaret Many Guns (may be North Piegan)	Aamsskáápipikani (Southern Piegan)	word:AT1969:1600	innáápiinowan	candy			N
AT1969	31	Tom Many Guns	Aamsskáápipikani (Southern Piegan)	word:AT1969:1941	apsáloka	Crow			N
AT1969	88	Tom Many Guns	Aamsskáápipikani (Southern Piegan)	word:AT1969:1604	innöösopaaʔtsis	bench (literally 'long seat')		Innoolsopaaʔtisi	N

Phase 2: Digitization

OriginalPartialWord	OriginalPartialWordTranslation	OriginalPartialWordCategory	OriginalPartialWordUR	OriginalPhrase	OriginalPhraseTranslation
	dance in the sun dance	AI+O -imaa- 'pseudo-intransitive on indefinite animate objects'; no related TA stem	-itapI:Skatimaa-		
				ánʔnuxkay kòoʔkúsi áaakitɛɛ••nʔixkyoʔpa	tonight we will sing and sing
	reflect, count		okaki		
	burn up	II -itii- 'with heat, accompanied by heat'	-Ittokiniti-		
	with, along, by	anaphoric preverb	i:/oSt		
	blood	Dep	-aaYapan-		
	be much	II -o- 'stative'	-a:kawo-		
	be much	AI -i(mm)- 'stative'	-a:kayI-		
nit-	my	pro. poss.			

Phase 3: Analysis of maximal stems, lemmas

WORD_ID	Word	Translation	LabStem	Stem_ID	LabStemCategory	LabLemma	Lemma_ID	LabLemmaTranslation	LabLemmaCategory
word:AT1969:0836	nàaáxsa	my grandparent	-àaáxs	stem:1095	NDA	-aaahs	1	grandparent	NDA
word:EC1911:006	ma-áħs	aunt, paternal (his grandmother)	-a-áħs	stem:3711	ND	-aaahs	1	grandparent	NDA
word:AT1969:1491	naaↃʔyi	my mouth	-aaↃ	stem:1750	NDI	-aaao	2	mouth	NDI
word:AT1969:1492	naaↃʔyi	my mouth	-aaↃ	stem:1751	NDI	-aaao	2	mouth	NDI
word:HH1885:045	n'ahaban	my blood	-'ahaban	stem:4239	ND	-aaapan	3	blood	ND
word:AT1969:1911	otáákiiʔsinʔa	womenfolk	-áákiiʔsin	stem:2170	NA	-aakii'sin	4	womenfolk	NA
word:AT1969:0837	nὲεέʔwa	my robe, my blanket	-ὲεέ	stem:1096	NDA	-aiai	5	robe	NDA

Current & Future Projects

Current & Future Projects

1. Synchronic morphophonology (current; tokenization)
   > Nisinoon project (Monica Macaulay and Hunter Lockwood)

2. Dialects and variation
   > "Documenting variation in Niitsi'powahsin" SSHRC Insight Grant
   > PIs: Inge Genee & Marie-Odile Junker

3. Historical change

4. Inflectional morphology and morphological parsers

Flexible structure, to accommodate future expansions?

- Adding fields: internal reconstruction, standardized glosses
- Adding sources: narratives, linguistic elicitation sessions
- etc.

Getting Started

- You don't need to know about databases
- Digitization takes the longest
- Many students and hours!
- Google Drive is handy (collaborating remotely and [a]synchronously)

Nitsíkohtaahsi'takihpinnaan!

Thank you! (It makes our heart happy.)

Special thanks to:

- Yale Spinup (esp. Tenyo Grozev!)
- GreenGeeks
- Blackfoot Lab members,
  incl. Alex Smith from ULeth!
- Danny Hieber
- Inge Genee
- Monica Macaulay
- Hunter Lockwood

References

Berman, Howard. 2006. Studies in Blackfoot prehistory. *IJAL* 72(2): 264–284.

Bliss, Heather. 2013. The Blackfoot configurationality conspiracy: Parallels and differences in clausal and nominal structures. University of British Columbia, PhD thesis.

Bliss, Heather and Elizabeth Ritter. 2009. "Speaker certainty, event realization and epistemic modality in Siksika Blackfoot. Ms, May 12, 2009.

Bloomfield, Leonard. 1946. Algonquian. In *Linguistic structures of Native America*, Hoijer, Harry (ed.), 85–129. (Publications in Anthropology 6). New York: Viking Fund.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2021. Ethnologue: Languages of the World. Twenty-fourth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.yale.idm.oclc.org. (Accessed: June 28, 2021.)

Frantz, Donald G. 2017. *Blackfoot grammar*. 3rd edn. University of Toronto Press.

Frantz, Donald G. 1978. Abstractness of phonology and Blackfoot orthography design. In A*pproaches to language, anthropological issues: Papers written for the IXth International Congress of Anthropological and Ethnological Sciences, Chicago, 1973*, McCormack, W. and S.A. Wurm (eds.), 307–325. Mouton Publishers.

References

Frantz, Donald G. and Norma Jean Russell. 2017. *Blackfoot dictionary of stems, roots, and affixes*. 3rd edn. University of Toronto Press.

Genee, Inge, and Marie-Odile Junker. 2018. The Blackfoot Language Resources and Digital Dictionary project: Creating integrated web resources for language documentation and revitalization. *LD&C* 12: 298–338.

Goddard, Ives. 2018. Blackfoot and Core Algonquian inflectional morphology: Archaisms and innovations. *PAC* 47: 83–106.

Goddard, Ives. 1994. The west-to-east cline in Algonquian dialectology. *PAC* 25: 187–211.

Isham, James. 1949/1743. *Observations on Hudsons Bay, 1743, and Notes and observations on a book entitled* A voyage to Hudsons Bay in the Dobbs Galley*, 1749*. Edited by E.E. Rich. (Publications of the Hudson's Bay Record Society 12.) Toronto: The Champlain Society.

Lacombe, Albert. 1886. *First reader in the English and Blackfoot languages, with pictures and words*. Montreal: C. O. Beauchemin & Son, Booksellers and Printers.

Lanning, Cassius. M. 1882. *A Grammar and Vocabulary of the Blackfoot Language*. Fort Benton, Montana Territory. Self-published.

References

Michelson, Truman. 1935. Phonetic shifts in Algonquian languages. *IJAL* 8(3/4): 131–171.

Peter, Naoki. 2014. Hiatus resolution in Blackfoot. Universität Bern, Masterarbeit.

Taylor, Allan. 1969. A Grammar of Blackfoot. Berkeley: University of California, Berkeley Dissertation.

Tims, John William. 1889. *Grammar and Dictionary of the Blackfoot Language in the Dominion of Canada*. London: Society for Promoting Christian Knowledge.

Weber, Natalie. 2021. Phase-based constraints within Match Theory. In Ryan Bennett et al. (eds.), *Supplemental Proceedings of the 2020 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America. (Accessed online: http://journals.linguisticsociety.org/proceedings/ index.php/amphonology/index.)

Weber, Natalie. 2020. *Syntax, prosody, and metrical structure in Blackfoot*. University of British Columbia, PhD dissertation.

Current & Future Projects

Synchronic morphophonology

- Nisinoon (https://nisinoon.net/)

Dialects and variation

- "there is as much variation between speakers from the same reserve as there is between speakers from different reserves" (Frantz and Russell, 2017: xiii)
- Documenting variation in Niitsi'powahsin (Blackfoot)
  [co-PIs: Inge Genee and Marie-Odile Junker]

Current & Future Projects

Historical change

- "Blackfoot is clearly the most **divergent** language in the Algonquian family" (Goddard:2018)
- Innovative sound changes "all contribute towards making Blackfoot vocabulary as a whole **appear as un-Algonquian**" (Michelson 1935: 142-143).
- "Indeed hitherto there has been a decided tendency to regard Blackfoot vocabulary as largely alien. I therefore state that nevertheless **the amount of Blackfoot vocabulary that can be shown to be of Algonquian origin is much greater than supposed** (I have several hundred etymologies which I think are certain), even if it is extremely difficult to enunciate phonetic shifts which "work" consistently' (Michelson 1935: 143).

Inflectional morphology

- currently not included
- many similarities with other Algonquian languages
- also some innovations and archaisms (Goddard 2018)
- add morphological parsers?