### *itwêwina* <u>https://itwewina.altlab.app</u>

**Towards a morphologically intelligent and user-friendly on-line dictionary of Plains Cree – next next round** Antti Arppe<sup>1</sup>, Eddie Antonio Santos<sup>1,2,</sup>, Andrew Neitsch<sup>1</sup>, Jolene Poulin<sup>1</sup>, Atticus Harrigan<sup>1</sup>, Katherine Schmirler<sup>1</sup>, Daniel Hieber<sup>1</sup> & Arok Wolvengrey<sup>3</sup> <sup>1</sup>Alberta Language Technology Lab, University of Alberta <sup>2</sup> Canadian Indigenous Languages Technology Project, National Research Council of Canada <sup>3</sup> First Nations University of Canada

### 21<sup>st</sup> Century Tools for Indigenous Languages Partnership Project

- <a href="http://altlab.artsrn.ualberta.ca">http://altlab.artsrn.ualberta.ca</a>
- https://21c.tools/
- Developing "low-hanging fruit" software for Indigenous languages
  - (1) intelligent electronic dictionaries (I-DICT)
  - (2) easily searchable, linguistically analyzed databases of spoken recordings and written texts
  - (3) spell-checkers (+ grammar-checkers), and
  - (4) intelligent computer-aided language learning applications (ICALL)
- Intelligent
  - make use of language technology
  - to deal with the complex word structure

### Acknowledgements

- Numerous native (Plains) Cree speakers and Elders
- Cree language instructors
  - − Miyo Wahkohtowin Education → Maskwacîs Education Schools Commission (MESC)
  - First Nations University of Canada
  - Wayne Jackson / nêhiyawêwin word of the day
- Long-standing, continuing, meticulous lexicographical work by Arok Wolvengrey
  - nêhiyawêwin : itwêwina / Cree : Words
- Funding through multiple grants from SSHRC, UAlberta, Kule Institute of Advanced Study, and support by National Research Council (NRC)
- Enduring encouragement
  - Arden Ogg (CLN), Earle Waugh (AECD), Ahmad Jawad (Intellimedia), Trond Trosterud, Sjur Moshagen, Lene Antonsen (Giellatekno/Divvun)

### Spoken dictionary of Maskwacîs Cree

Mary Jean Littlechild, Louise Wildcat, Jerry Roasting, Harley Simon, Annette Lee, Arlene Makinaw, Rosie Rowan, Rose Makinaw, Kisikaw, Betty Simon, Brian Lightning, Brian Lee, Linda Oldpan, Mariam Buffalo, Debora Young, Ivy Raine, Paula Makinaw, Norma Linda Saddleback, Renee Makinaw, Atticus Harrigan, Katherine Schmirler, Dustin Bowers, Megan Bontohon, Sarah Giesbrecht, Patricia Johnson, Timothy Mills, Jordan Lachler & Antti Arppe

Maskwacîs Education Schools Commission (MESC)

#### Morphological intelligence – rich word structure (English vs. Plains Cree)

- see
  - sees
  - saw
  - seen
  - seeing
  - I see  $\not{O} \rightarrow$  I see it  $\rightarrow$  I see him/her

- wâpiw 's/he sees Ø'
- wâpahtam 's/he sees it'
- wâpamêw 's/he see him/her'
  - kiwâpamin 'You [one] see me'
  - *ê-wâpamiyan* 'You [one] see me'
  - kiwâpaminân 'You [one] see us'
  - *ê-wâpamiyâhk* 'You [one] see us'
  - kiwâpaminâwâw 'You all see me'
  - ê-wâpamiyêk 'You all see me'
  - kiwâpaminân 'You all see us'
  - *ê-wâpamiyâhk* 'You all see us
  - and up to 500 other core forms
  - plus thousands more complex forms (with 1-N preverbs)
- pê-wâpamêw

...

• nikî-nitawi-wâpamâw

## Community/User feedback → initial design decisions

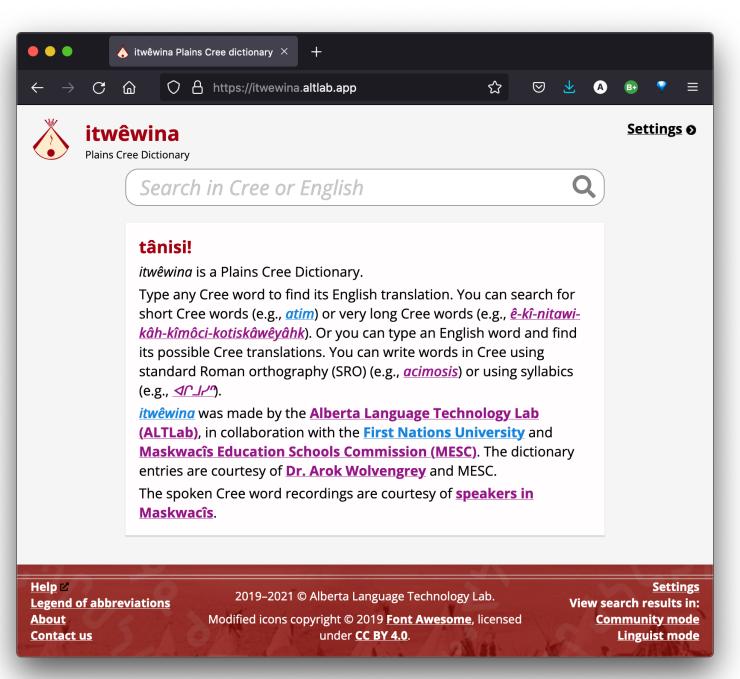
- ← Language instructors ← feedback sessions, CILLDI courses delivered in 2020, feedback emails
- $\rightarrow$  simple, but extendible
- Automatic handling of Cree/English as search language
- Showing only most essential core information first
   But full lexicographical and other details easily accessible
- Plain (English) terminology as primary
  - But linguistic terminology/analysis easily accessible

## Meeting expectations – last year → current status?

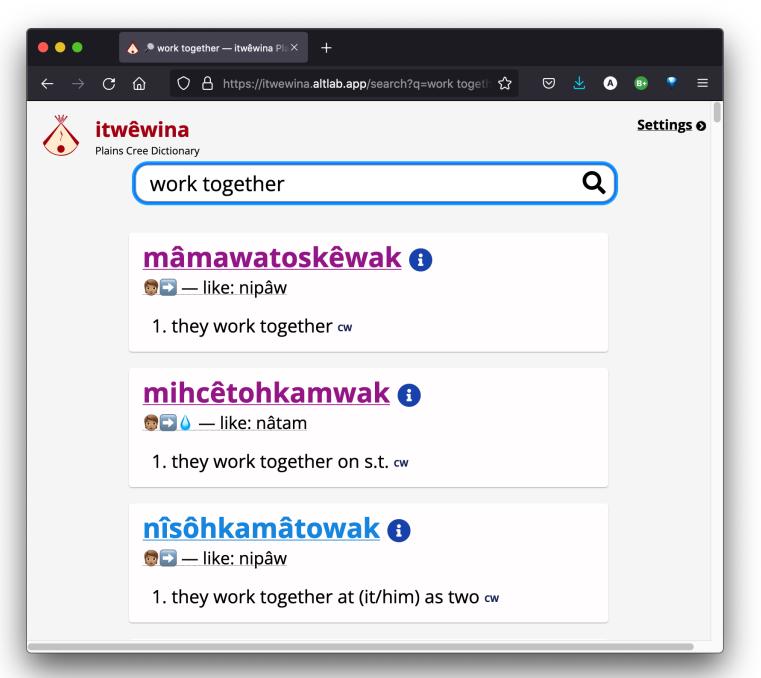
- Facebook: nêhiyawêwin word of the day what linguistic help do people ask for?
  - Cree naming
    - What is the Cree word for X (English concept)?
    - → Good start → multiword concepts
  - Cree phrases
    - How do I say X (English phrase) in Cree?
    - → Step-wise through looking up paradigms → direct linking of English phrases to matching Cree word forms
  - Cree syllabics
    - How do I write X (Cree word) in Cree syllabics?
    - $\rightarrow$  Done
  - Cree pronunciation
    - How do I say X (Cree word) in Cree?
    - $\rightarrow$  Available for a part of the dictionary entries  $\rightarrow$  increasing coverage to cover much more of the Cree paradigms
  - Cree-to-English translation
    - What does X (Cree word) mean in English?
    - → good start → improving fuzzy matching
- Devices/Contexts
  - Most users have majority language keyboards  $\rightarrow$  diacritic characters present a challenge
  - Mobile gadgets, rather than desktops or even laptops
    - $\rightarrow$  Designing for the smaller screens
    - ightarrow mobile, non-physical keyboards ightarrow typing of characters even more fuzzy
  - Internet connectivity is not guaranteed ← in particular in remote communities

### What is the Cree word for X?

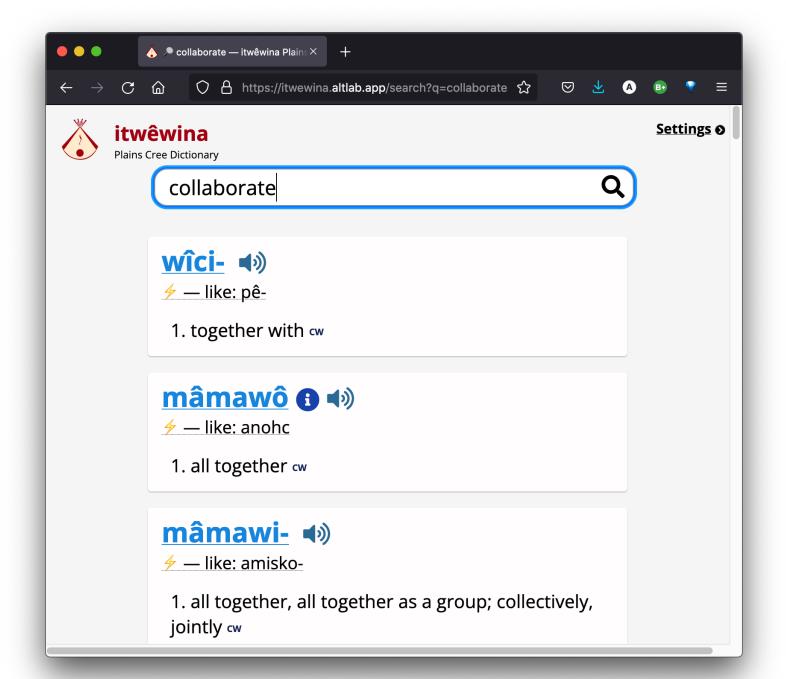
- multi-word searches
- "missing" rare words
- synonyms



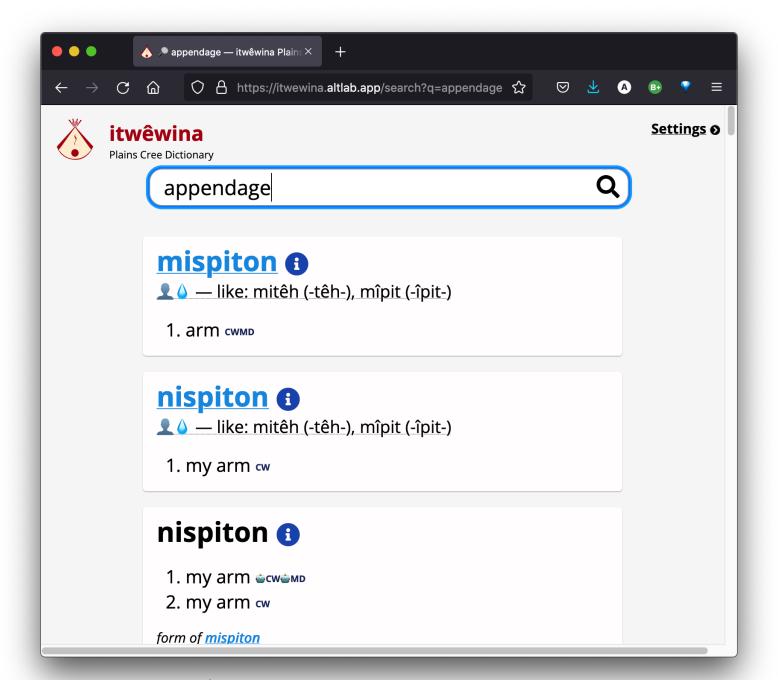
itwêwina landing page, for initiating searches with Cree/English words



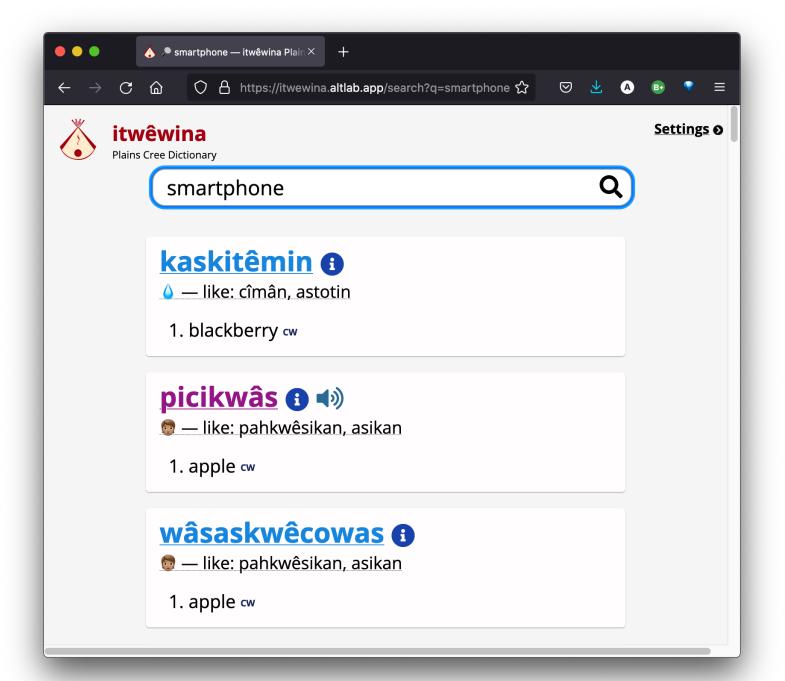
Searching with a multi-word expression: *work together* 



Searching with a rarer/synonymous word not in the English definitions: collaborate



Searching with a rarer/synonymous word not in the English definitions: *appendage* 



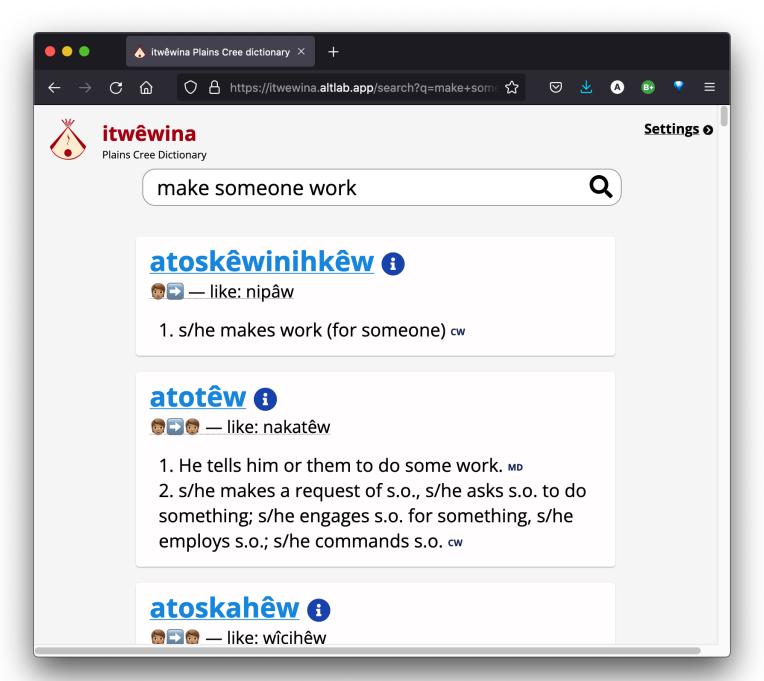
Searching with a rarer/synonymous word not in the English definitions: *smartphone* 

## How was this improvement to search achieved?

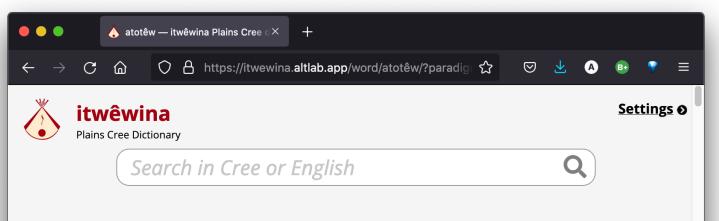
- Using multiple criteria to rank the relevance of the search results
  - edit-distance
  - corpus—based frequency
  - morpheme frequency
  - survey of relevance ranking of search results
  - semantic vectors (trained on huge English corpora
    - ← Google News Corpus)
      - applied to search words, and
      - English definitions

### How do I say X in Cree?

 Matching English phrases directly with most frequent/central Cree word-forms



Searching with an English multi-word concept: make (someone) work



#### atotêw

👦 🖸 🧶 — like: nakatêw

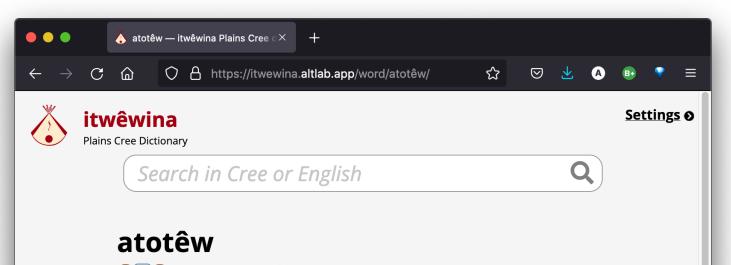
VTA-4: transitive animate verb – class 4

 He tells him or them to do some work. MD
 s/he makes a request of s.o., s/he asks s.o. to do something; s/he engages s.o. for something, s/he employs s.o.; s/he commands s.o. cw

► Show plain English labels

	something is happening	g now	
	ni-/ki- word	ê-/kâ- word	
$I \rightarrow you$ (one)	kitatotitin	ê-atotitân	
$I \rightarrow him/her$	nitatotâw	ê-atotak	
$I \rightarrow you$ (all)	kitatotitinâwâw	ê-atotitakok	

Looking up an inflected word-form in the paradigm: *I made them work* 



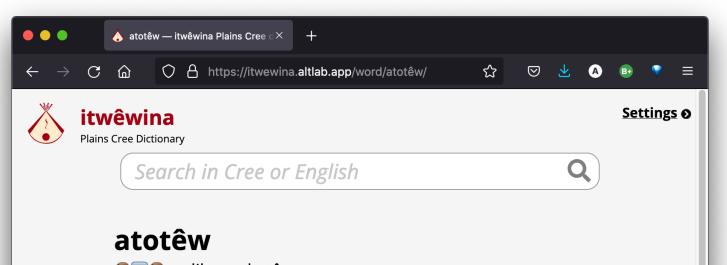
👦 🖸 🧶 — like: nakatêw

VTA-4: transitive animate verb – class 4

 He tells him or them to do some work. MD
 s/he makes a request of s.o., s/he asks s.o. to do something; s/he engages s.o. for something, s/he employs s.o.; s/he commands s.o. cw

IndPlain English labels $1s \rightarrow 2s$ kita $1s \rightarrow 3s$ nita $2s \rightarrow 1s$ kita $2s \rightarrow 1s$ kita $2s \rightarrow 3s$ kita

Changing the paradigm labeling to scientific / linguistic



👦 🖸 🛑 like: nakatêw

VTA-4: transitive animate verb – class 4

 He tells him or them to do some work. MD
 s/he makes a request of s.o., s/he asks s.o. to do something; s/he engages s.o. for something, s/he employs s.o.; s/he commands s.o. cw



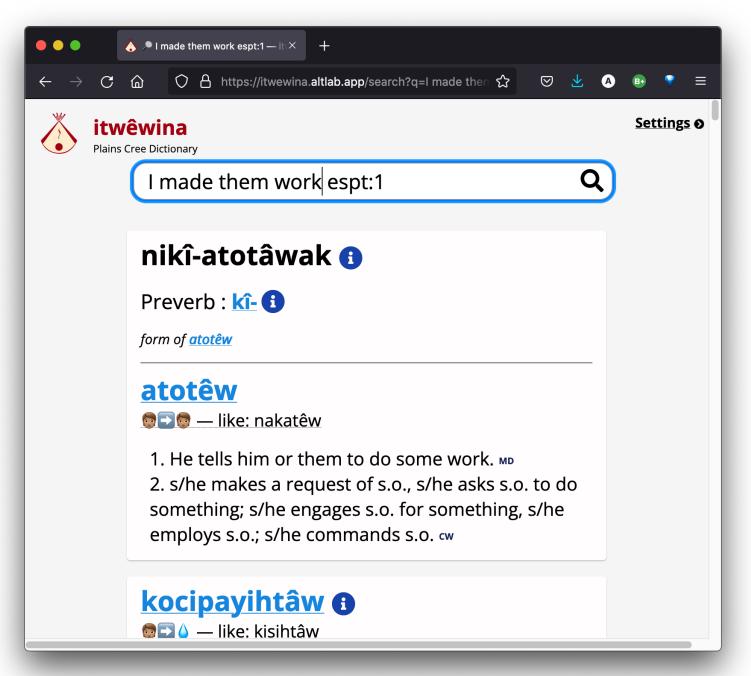
Changing the paradigm labeling to plain *nêhiyawêwin* 

	otêw — itwêwina Plains Cree	× +					
$\leftarrow$ $\rightarrow$ C @	○ A https://itwe	wina.altlab.app/word/atotê	w/?paradig ☆ ♡	¥ (	A 📴	Ŷ	
		something happened ea	rlier				
		ni-/ki- word	ê-/kâ- word				
	$I \rightarrow you$ (one)	kikî-at <mark>ptitin</mark>	ê-kî-atotitân				
	$I \rightarrow him/her$	nikî-at <mark>otâw</mark>	ê-kî-atotak				
	$I \rightarrow you$ (all)	kikî-at <mark>p</mark> titinâwâw	ê-kî-atotitakok				
	$I \rightarrow them$	nikî-atotâwak	ê-kî-atotakik				
	$I \rightarrow$	nikî-atotimâwa	ê-kî-atotimak				
	him/her/them						
		ni-/ki- word	ê-/kâ- word				
	you (one) $ ightarrow$ me	kikî-atosin	ê-kî-atosiyan				
	you (one) → him/her	kikî-atotâw	ê-kî-atotat				
	you (one) → we (but not you)	kikî-atosinân	ê-kî-atosiyâhk				
	you (one) $ ightarrow$ them	kikî-atotâwak	ê-kî-atotacik				
	you (one) → him/her/them	kikî-atotimâwa	ê-kî-atotimat				
		ni-/ki- word	ê-/kâ- word				
	we (but not	kikî-atotitinân	ê-kî-atotitâhk				

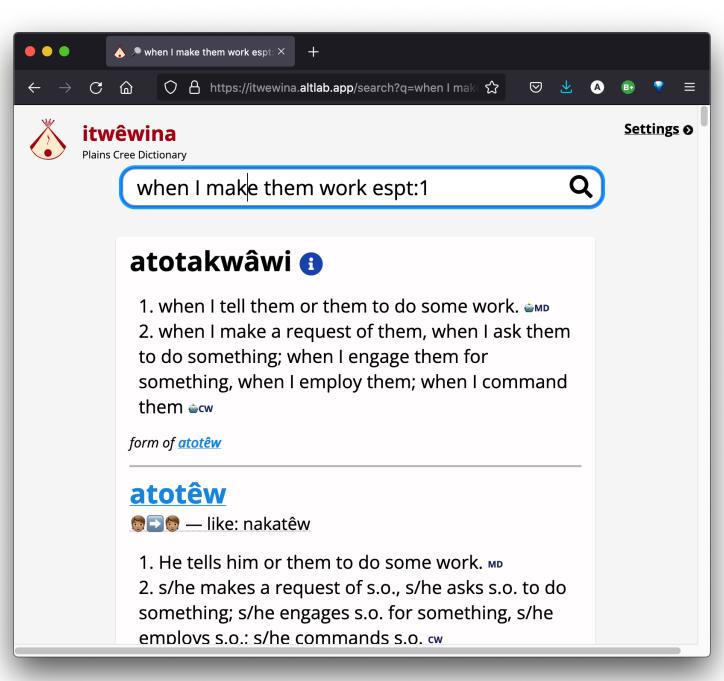
Looking up an inflected word-form in the paradigm: Prt+1Sg+3PlO → I made them work

### How do I say X in Cree?

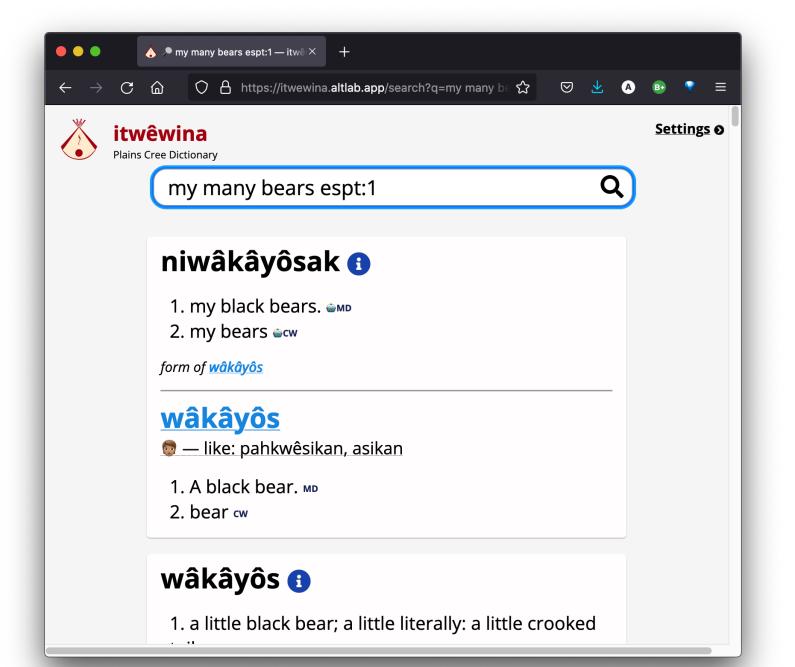
- What is the Cree word-form that corresponds to a simple English phrase?
- Better solution that paradigm lookup?



Looking up with a simple English phrase: *I made them work* 



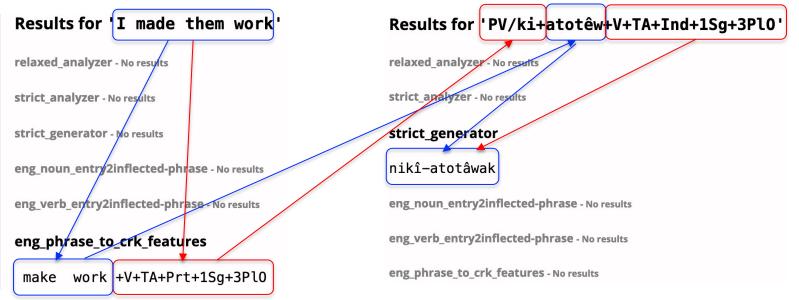
Looking up with another simple English phrase: when I make them work



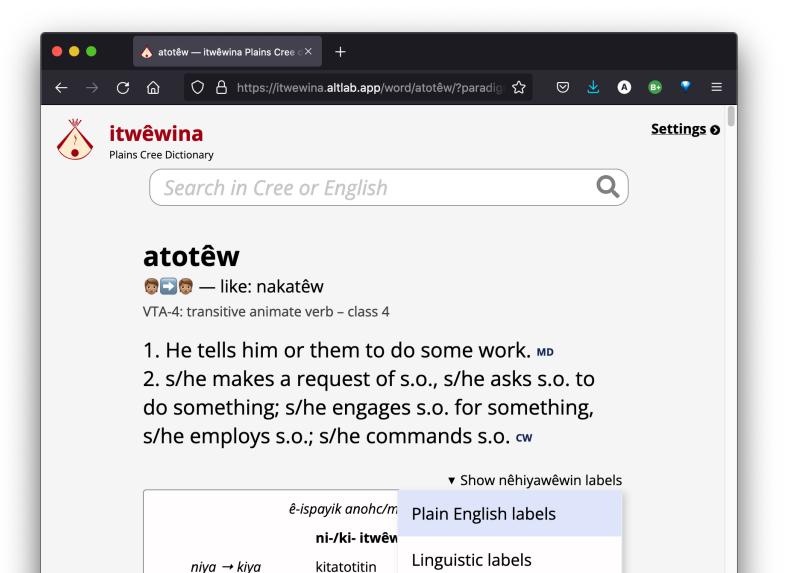
Looking up with a simple English phrase: my many bears

### How is this achieved?

- A computational model for the analysis of simple English phrases
- These are matched with the Cree word-forms in the full paradigms



# How do I write X in Cree syllabics?



Choosing Western Cree syllabics as the display orthography for the Cree content

Nêhiyawêwin labels

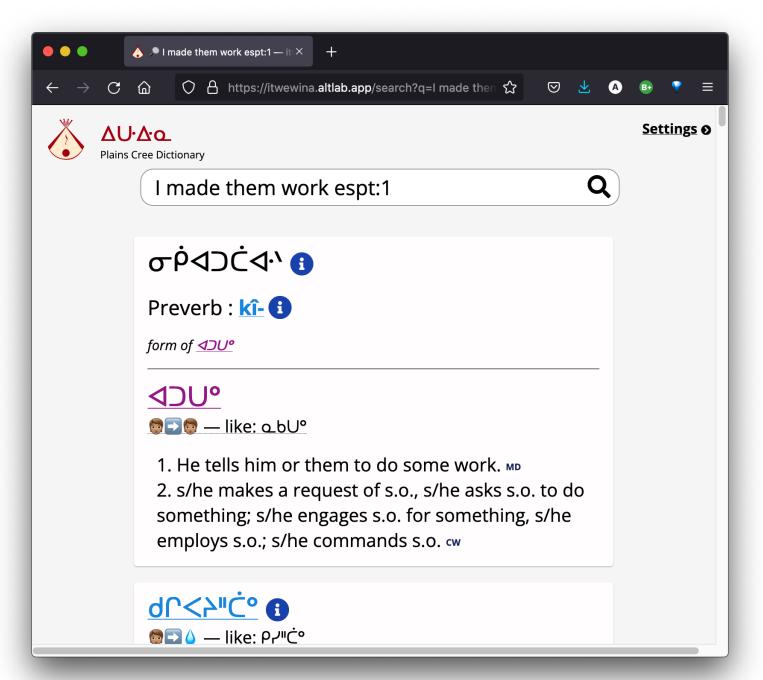
nitatotâw

kitatotitinâwa

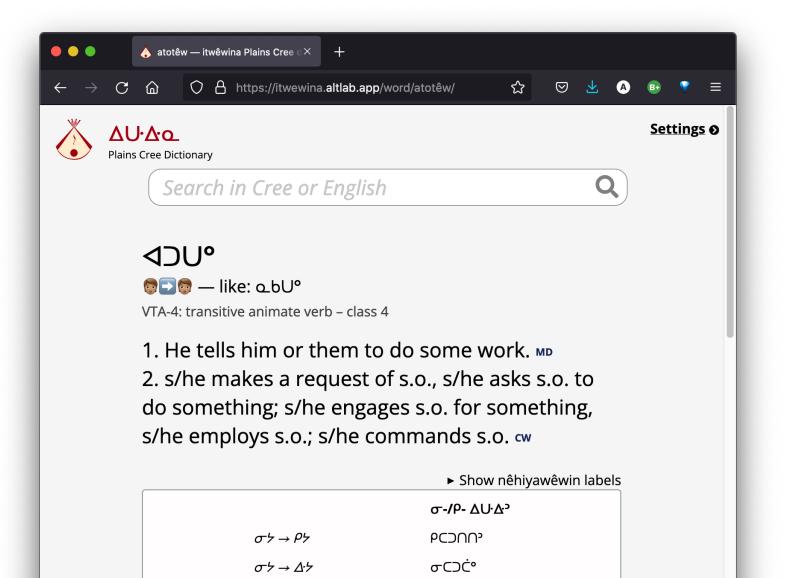
niya → wiya

niya →

1.1. . . . . . . . .



Choosing Western Cree syllabics as the orthography of the Cree content



Choosing *Western Cree syllabics* as the orthography of the Cree content – also in the paradigms

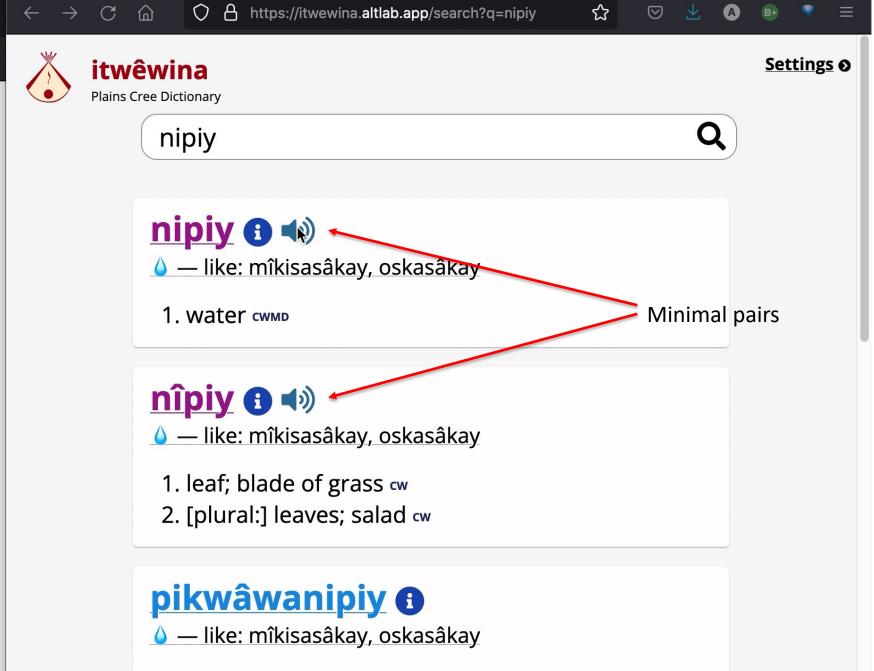
ירכס

PCDĊ°

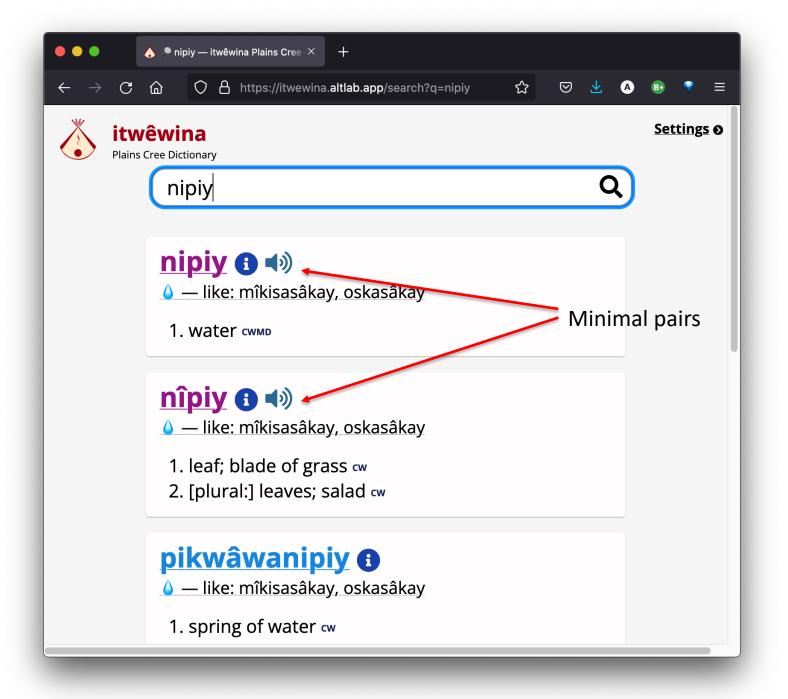
 $\rho \rightarrow \sigma \rightarrow \sigma$ 

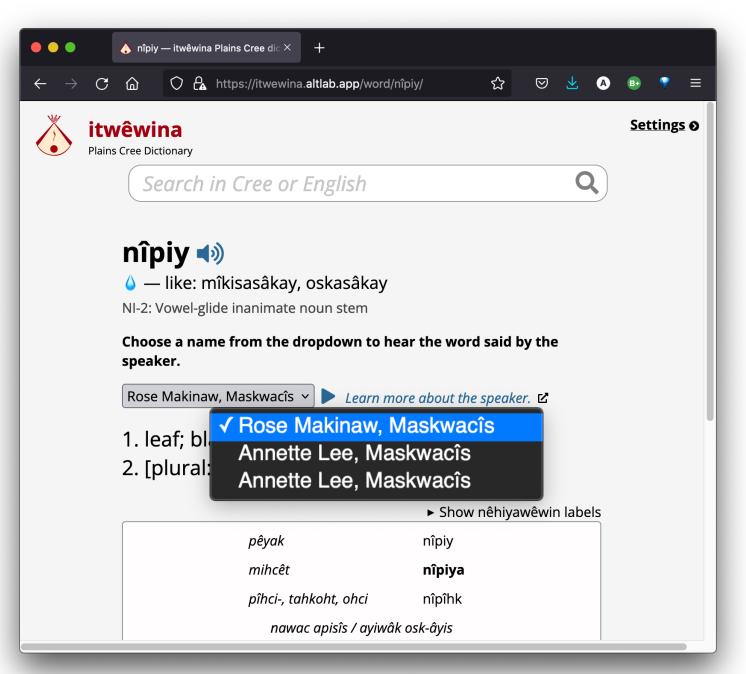
 $\rho \rightarrow \Delta \dot{\gamma}$ 

#### How do I **SAY** X in Cree?

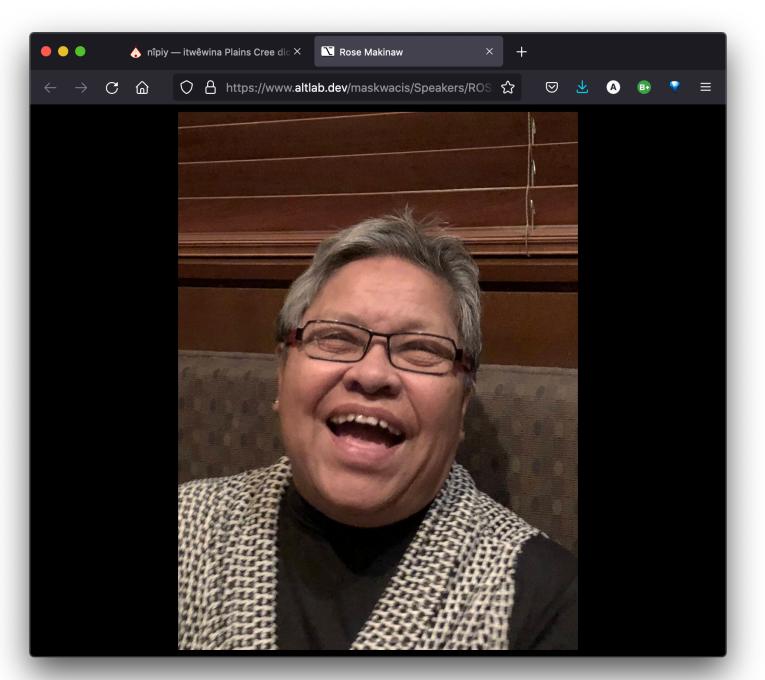


1. spring of water w

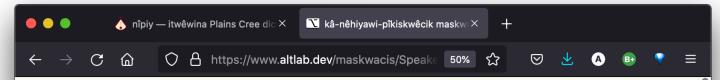




Selecting another speaker for their recording



Looking up the speaker page for Rose Makinaw (Maskwacîs)



#### kâ-nêhiyawi-pîkiskwêcik maskwacîsihk — Maskwacîs Cree Speakers











Mary Jane Littlechild

Louise Wildcat

Jerry Roasting



Arlene Makinaw





Rose Makinaw

Kisikaw



Betty Simon



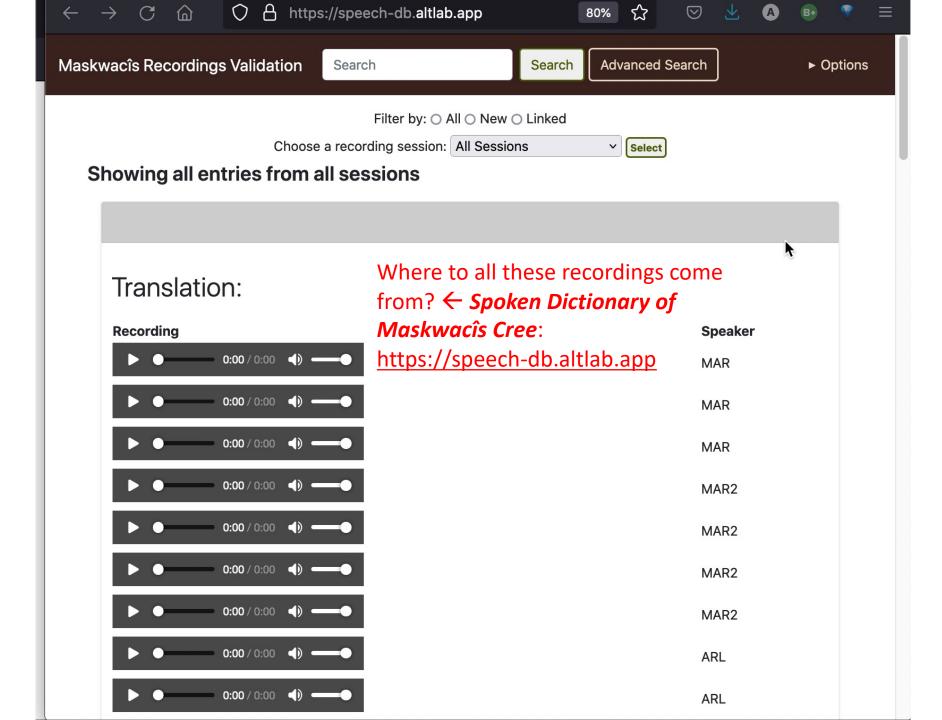


Rosie Rowan



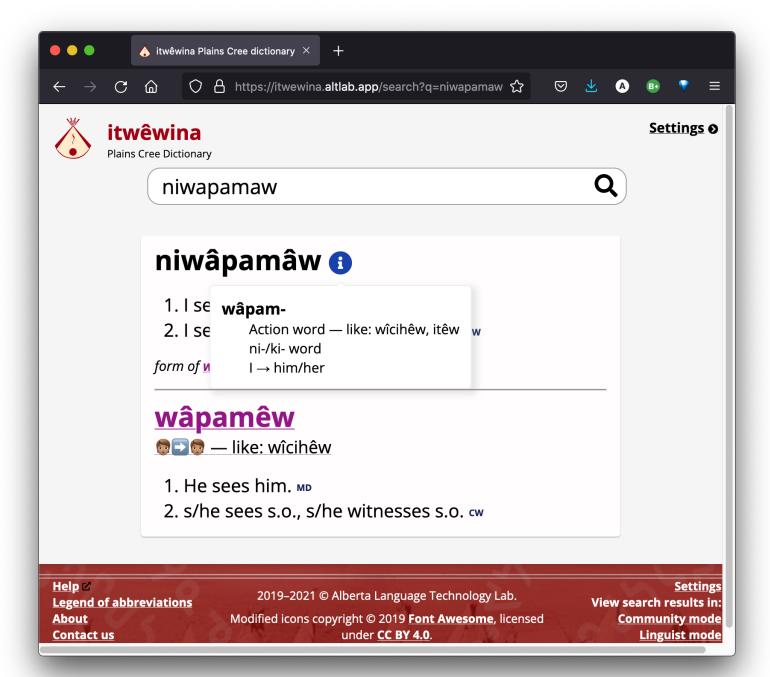




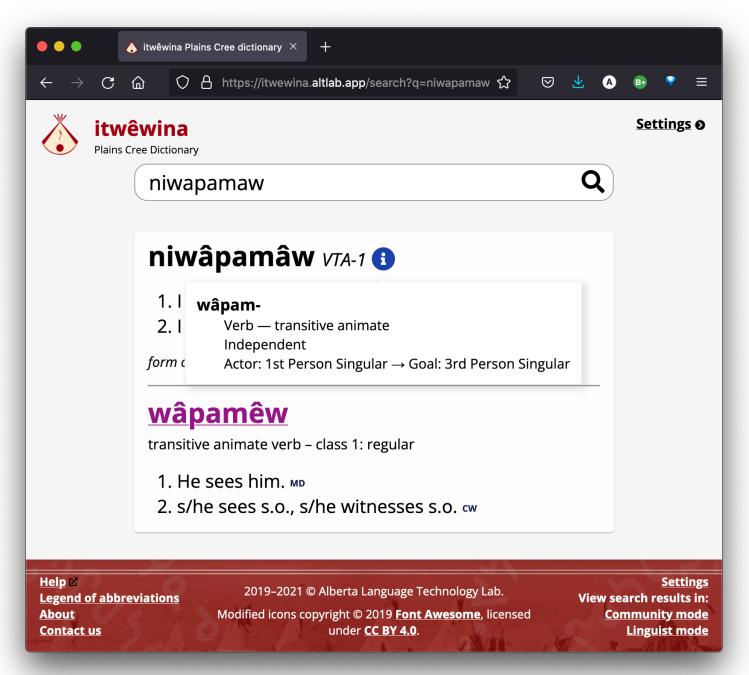


# What does X (Cree) mean (in English)?

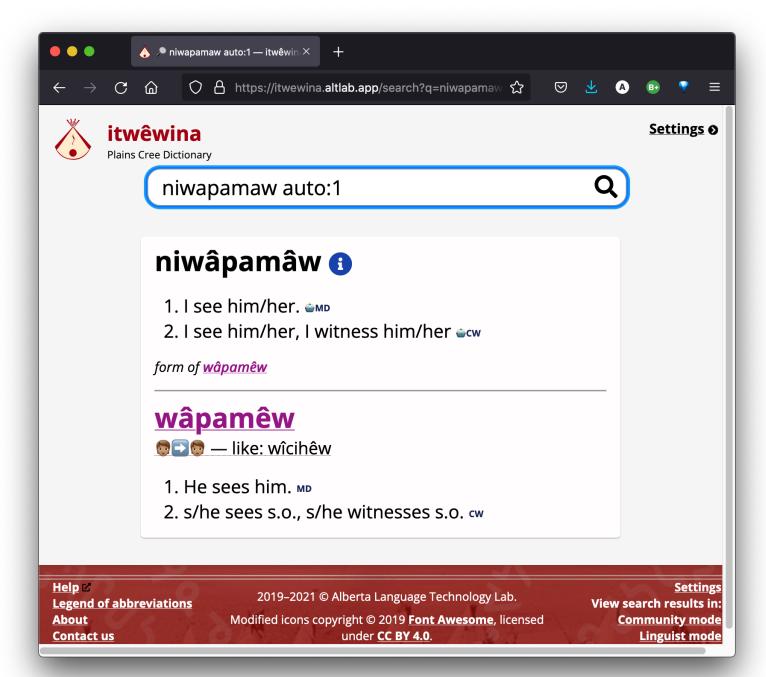
 Providing a simple English phrase as a translation, rather than a linguistic analysis



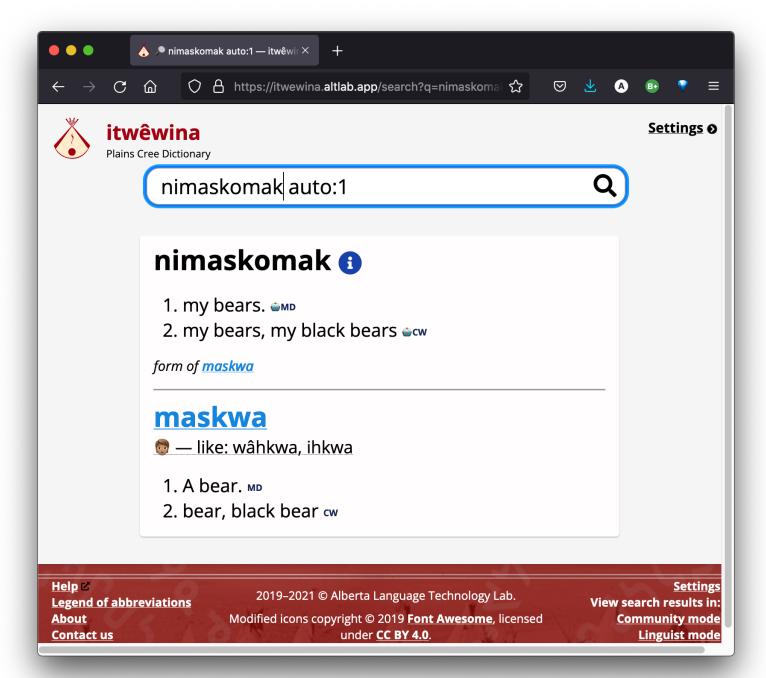
Looking up the relatively common Cree word-form: *niwâpamâw -> analysis in pop-up* 



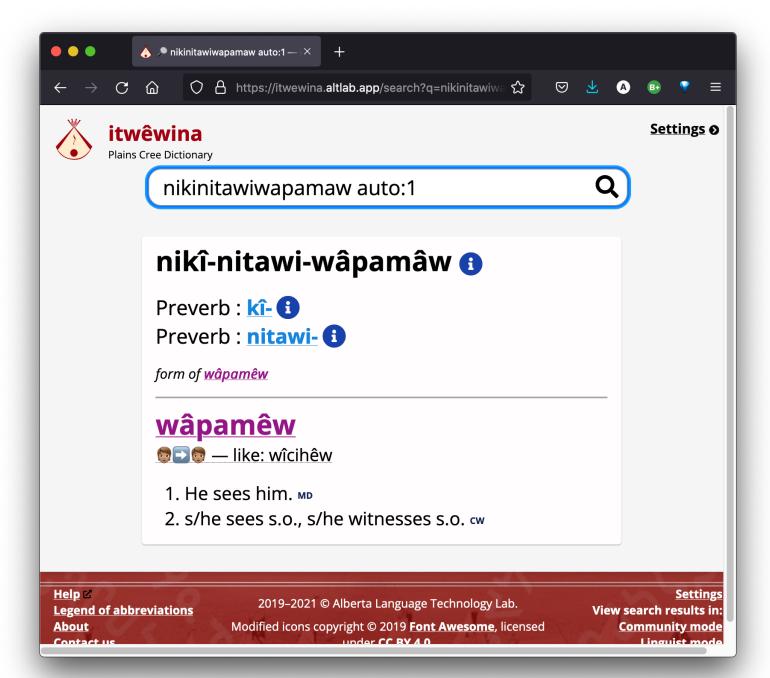
Swapping the labeling of linguistic features to: *linguistic labels* 



Looking up the relatively common Cree word-form: *niwâpamâw > I see him/her, ...* 



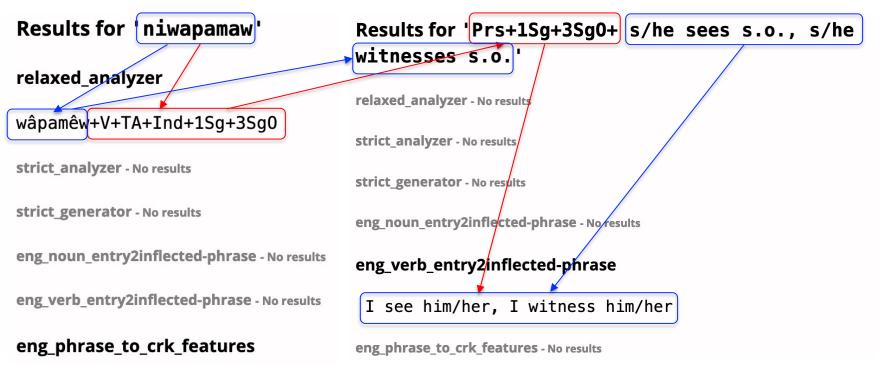
Looking up the relatively common Cree word-form: *nimaskomak* -> my bears, ...



For more complex Cree word-forms: *nikî-nitawi-wâpamâw → only pop-up analysis available* 

### How is this achieved?

 Linguistic analyses for Cree word-forms are used to computationally generate corresponding simple English phrases



#### Meeting expectations – current status?

- Cree naming
  - What is the Cree word for X (English concept)?
  - → multiword concepts, rare words and synonyms now provide good matches
- Cree phrase
  - How do I say X (English phrase) in Cree?

→ Step-wise, through looking up paradigms → simple English phrases are directly linked to matching Cree word forms → increasing phrase coverage

- Cree syllabics
  - How do I write X (Cree word) in Cree syllabics?
  - → Done
- Cree pronunciation
  - How do I say X (Cree word) in Cree?

 $\rightarrow$  more and more recordings are being gradually added  $\rightarrow$  continuing to add more recordings + synthesizing word-forms that have not been recorded

- Cree-to-English translation
  - What does X (Cree word) mean in English?

→ plain English representation of linguistic analyses → rough English translations can be provided for the most frequent inflected Cree word forms → increasing phrase coverage

# Other possible/necessary developments?

- Mobile/off-line version of itwêwina available for beta testing
  - minimal set of functionalities
  - BUT can be used without an Internet/WiFi connection
- Full functionalities available in **on-line** *itwêwina*

# Other possible/necessary developments?

- On-line dictionary as a language learning/instruction tool
  - Cree word building
    - What are little Cree words made up of?
    - Showing morpheme boundaries, in paradigms and throughout (inflection → derivation)
  - Linguistic reference data  $\rightarrow$  "grammar" pages
- Meeting the needs and expectations of multiple user communities
  - Choice of linguistic terminology  $\rightarrow$  linguistic analysis, paradigms
  - − Choice of linguistic interface → Plain English vs. Plain nêhiyawêwin vs. Linguistic
  - Semantic classification of dictionary entries (already exists)

### Morphological breakdown ← already supported by computational model

	One	Many	Further	In/On
someone's	_	_	_	_
my	n·ôhkom	n <i>·</i> ôhkom∙ak	n·ôhkom·a	n∙ôhkom∙ihk
your (one)	k·ôhkom	k·ôhkom·ak	k·ôhkom∙a	k∙ôhkom∙ihk
his/her			o∙hkom∙a	ohkom·ihk
our (but not your)	n∙ôhkom∙inân	n·ôhkom·inân·ak	n·ôhkom·inân·a	n∙ôhkom∙inâ∙hk
your and our	k·ôhkom∙inaw	k·ôhkom·inaw·ak	k·ôhkom·inaw·a	k∙ôhkom∙inâ∙hk
your (all)	k∙ôhkom∙iwâw	k·ôhkom·iwâw·ak	k ôhkom iwâw a	k∙ôhkom∙iw∙âhk
their			ohkom∙iwâw∙a	ohkom∙iw∙âhk
his/her/their (further)			ohkom∙iyiw∙a	ohkom∙iy∙ihk

Inflectional morpheme boundaries shown in the paradigm of *nôhkom* 'my grandmother'

#### Spoken Cree paradigms $\leftarrow$ synthesizer already exists

🔴 🔴 🌒 👗 🖉 v	viyapamat — itwêwina: the X	itwêwina - Online Plains Cree X	Cree-intelligent-dictionary/cri	Cree Audio Synthesis X
$\leftrightarrow$ > C $\textcircled{a}$	🗊 🔏 paradigm-sou	und-thingy.s3-website.us-east	-2.amazona 🧿%   🚥 🖂 🏠	¥ III 🖾 🛛 🖸
Present			Future Intentional	
Person	Independent	Conjunct	Independent	
1Sg→3SgO	niwâpamâw 🕠	ê-wâpamak	nika- 🌒 wâpamâw	
2Sg→3SgO	kiwâpamâw 📢)	ê-wâpamat	<b>▲)</b> kika- <b>▲)</b> wâpamâw	
1PI→3SgO	niwâpamânân I	●) ê-wâpamâyâhk ■♪	nika- wâpamânân	<b>■</b> »
12PI→3SgO	kiwâpamânaw	●) ê-wâpamâyahk ■♪	kika- wâpamânaw	<b>■</b> »
2PI→3SgO	kiwâpamâwâw	ê-wâpamâyêk	kika- wâpamâwâw	<b>■</b> >
1Sg→3PIO	niwâpamâwak	ê-wâpamakik	nika- wâpamâwak	<b>■</b> »
2Sg→3PlO	kiwâpamâwak 🛛	ê-wâpamacik	kika- wâpamâwak	<b>■</b> »
1PI→3PIO	niwâpamânânak	ê-wâpamâyâhkik	nika- wâpamânânak	<b>■</b> >

Synthesized recordings shown for full paradigm of: *wâpamêw* 

## Final words

 Developing an on-line language-technological application for an Indigenous language community – that the community will fully use – is much more than a one-off delivery

← gathering feedback and adjusting the tool accordingly requires time

→ feedback can change over time, as experience with the tool grows

 Software coding development is the bottleneck, not so much the creation of linguistic models and resources

#### Thank you – ay hay – kiitos!

#### Questions? Comments! Feedback?

Application: <u>https://itwewina.altlab.app/</u> Instructions: <u>https://altlab.artsrn.ualberta.ca/itwewina/</u> Source code: <u>https://github.com/UAlbertaALTLab/morphodict/</u>