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

Towards a morphologically intelligent and user-friendly on-line dictionary of Plains Cree – next next round

Antti Arppe¹, Jolene Poulin¹, Eddie Antonio Santos^{1,2}, Andrew Neitsch¹, Atticus Harrigan¹, Katherine Schmirler¹, Daniel Hieber¹, Ansh Dubey¹ & Arok Wolvengrey³

 Alberta Language Technology Lab, University of Alberta
 Canadian Indigenous Languages Technology Project, National Research Council of Canada
 First Nations University of Canada

21st Century Tools for Indigenous Languages Partnership Project

- http://altlab.artsrn.ualberta.ca
- 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) intelligent spell-checkers (+ grammar-checkers) and keyboards, 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, and Kule Institute of Advanced Study
- 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 Bontogon, 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 Ø□ I see it □ 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 's/he comes and sees him/her
- nikî-nitawi-wâpamâw 'I went and saw him/her

. . .

Community/User feedback → initial design decisions

- ← Language instructors, feedback sessions, CILLDI courses delivered in 2020, feedback emails
- → simple, but extendible
- Automatic handling of both Cree and 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 – Fall 2020 \rightarrow Fall 2021 \rightarrow Fall 2022

- 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?
 - Done
 - Cree pronunciation
 - How do I say X (Cree word) in Cree?
 - Available for a part of the dictionary entries → increasing coverage of the Cree entries →
 adding fully spoken paradigms for select entries ← native speakers + synthesized
 - Cree-to-English translation
 - What does X (Cree word) mean in English?
 - good start → improving fuzzy matching

Meeting expectations – Fall 2020 \rightarrow Fall 2021 \rightarrow Fall 2022

- On-line dictionary as a language learning/instruction tool
 - Cree word building
 - What are little Cree words made up of?
 - Showing inflectional morpheme boundaries, in paradigms and throughout
- Meeting the needs and expectations of multiple user communities
 - (Spoken paradigms ← spoken examples are highly valued by Cree communities)
 - Community identity
 - user can select dictionary or speech community (→ spoken examples)
 - Cree-to-English content added from Alberta Elders' Cree Dictionary ← highly valued by Cree communities in Alberta
 - Fully implemented optionality for different user groups
 - user can select preferred format for any feature with multiple available options

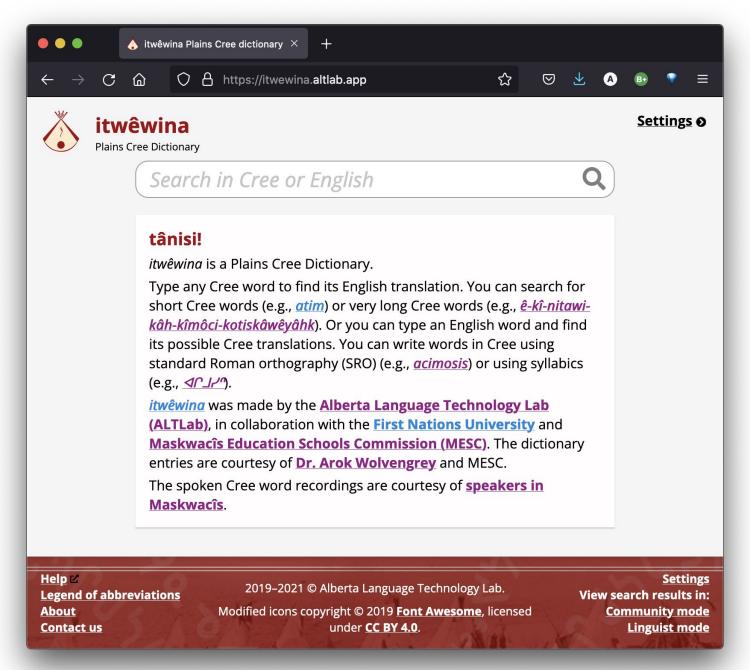
Meeting expectations – Fall 2020 \rightarrow Fall 2021 \rightarrow Fall 2022

- Devices/Contexts
 - Most users have majority language keyboards → diacritic characters present a challenge
 - Mobile gadgets, rather than desktops or even laptops
 - → Designing for the smaller screens
 - \rightarrow mobile, non-physical keyboards \rightarrow typing of characters even more fuzzy
 - Internet connectivity is not guaranteed ← in particular in remote communities
 - Most users in Indigenous language communities have
 Android rather than iOS mobile devices

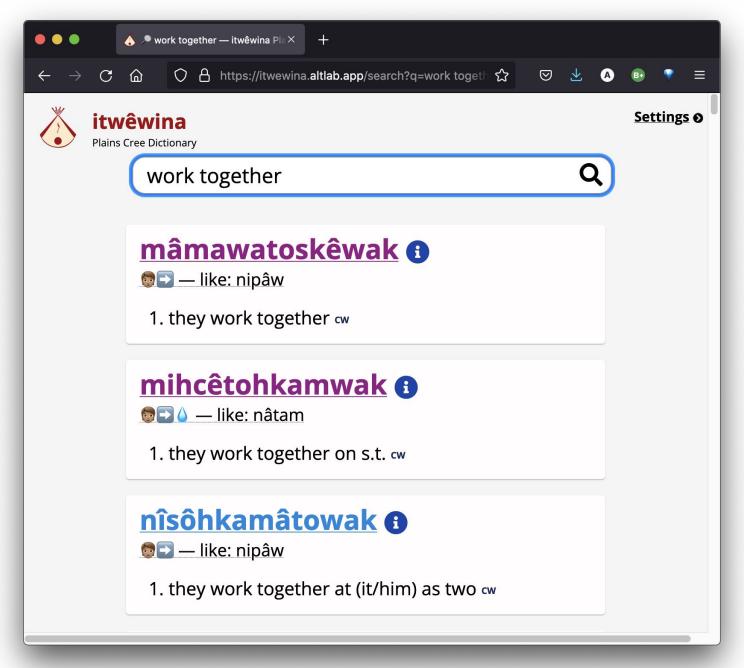
Developments up to Fall 2022

What is the Cree word for X?

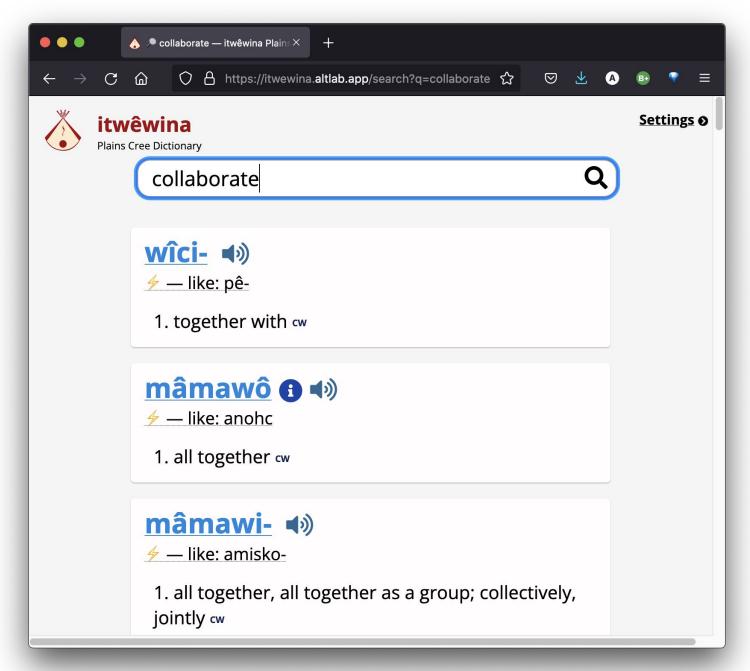
- multi-word searches
- "missing" rare words
- synonyms



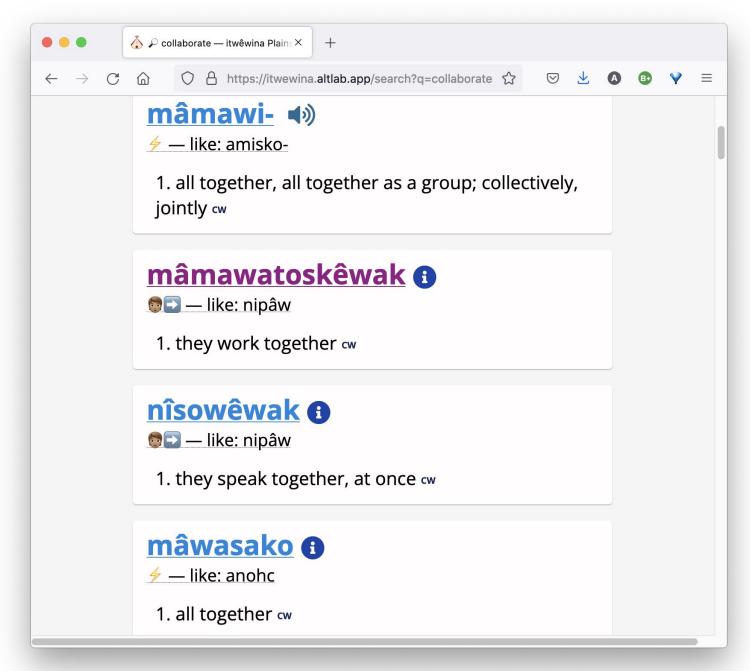
itwêwina landing page, for initiating searches with Cree/English words



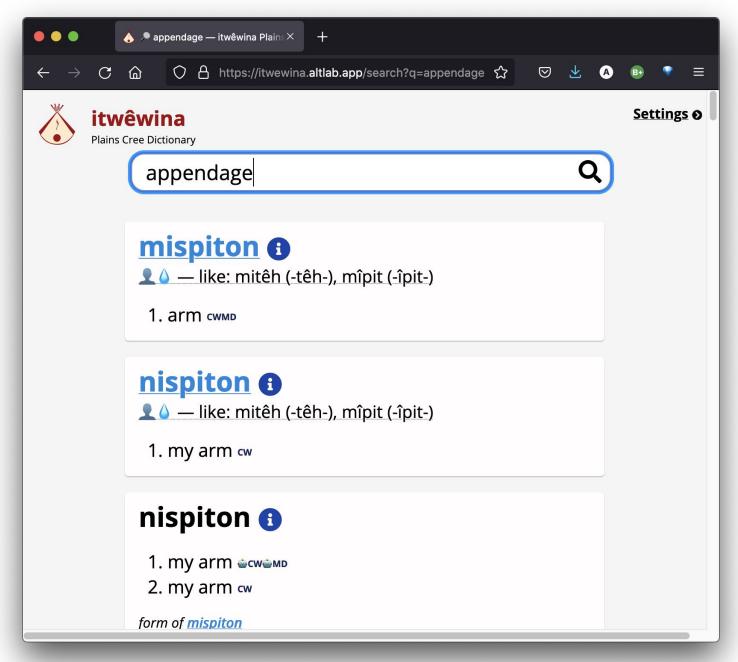
Searching with a multi-word expression: work together



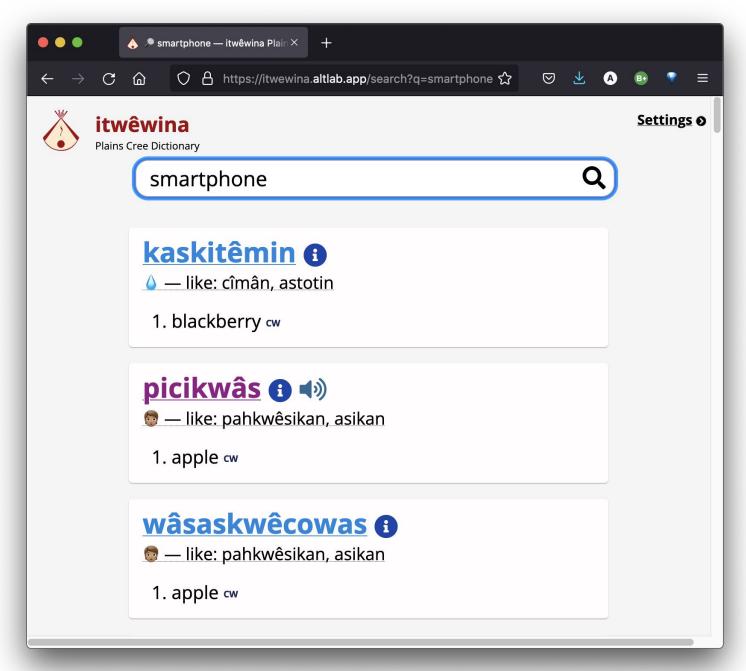
Searching with a rarer/synonymous word not in the English definitions: collaborate



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



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



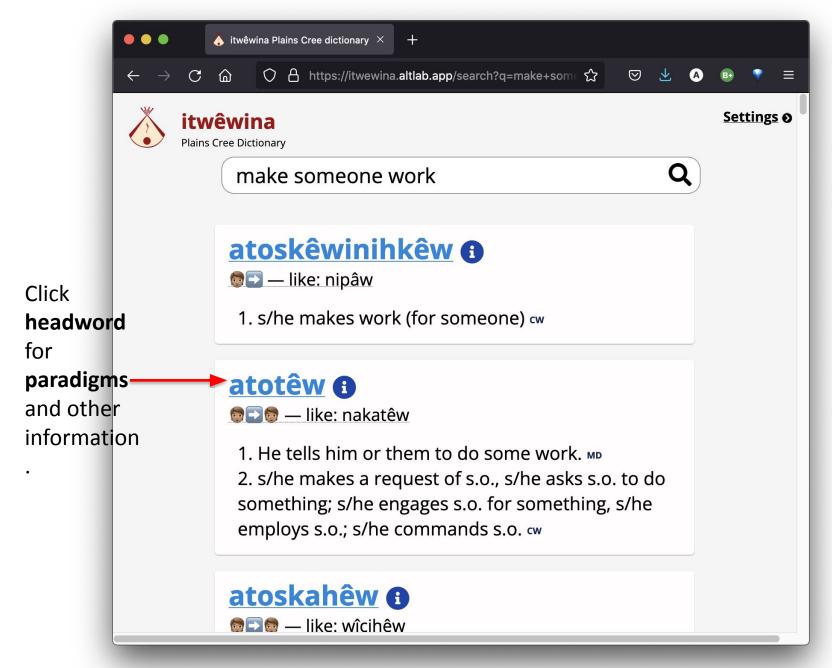
Searching with a rarer/synonymous word not in the English definitions: smartphone

How was this improvement to search achieved?

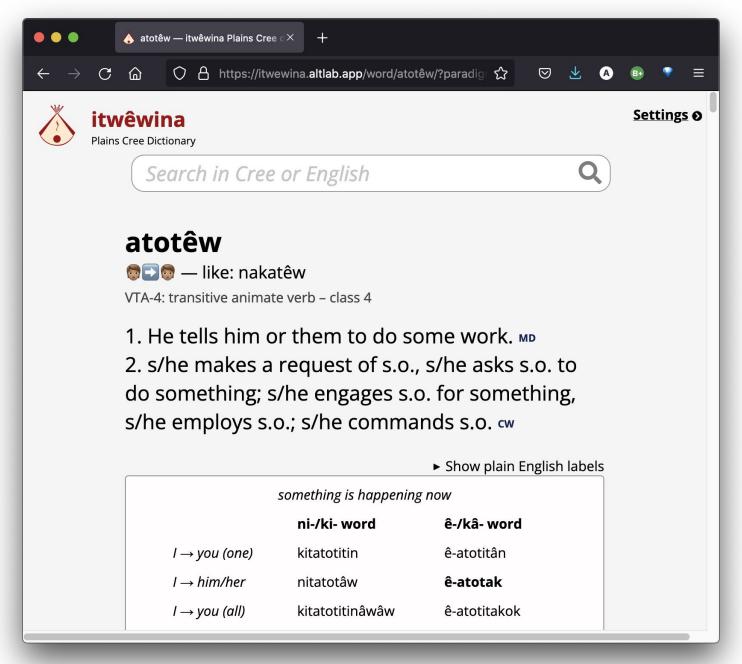
- Using multiple criteria to rank the relevance of the search results
 - edit-distance
 - occurrence in glossaries of introductory Cree textbooks
 - corpus-based frequency
 - morpheme frequency
 - user survey of relevance ranking of search results
- semantic vectors (trained on huge English corpora with billions of words ← Google News Corpus)
 - applied to search words, and
 - English definitions

How do I say X in Cree? — Part 1

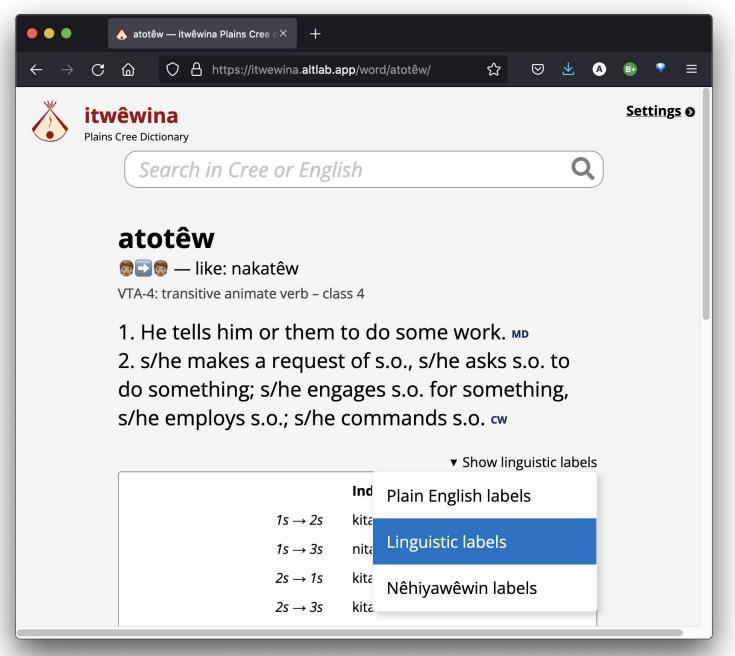
 Matching simple English phrases directly with most frequent/central Cree word-forms



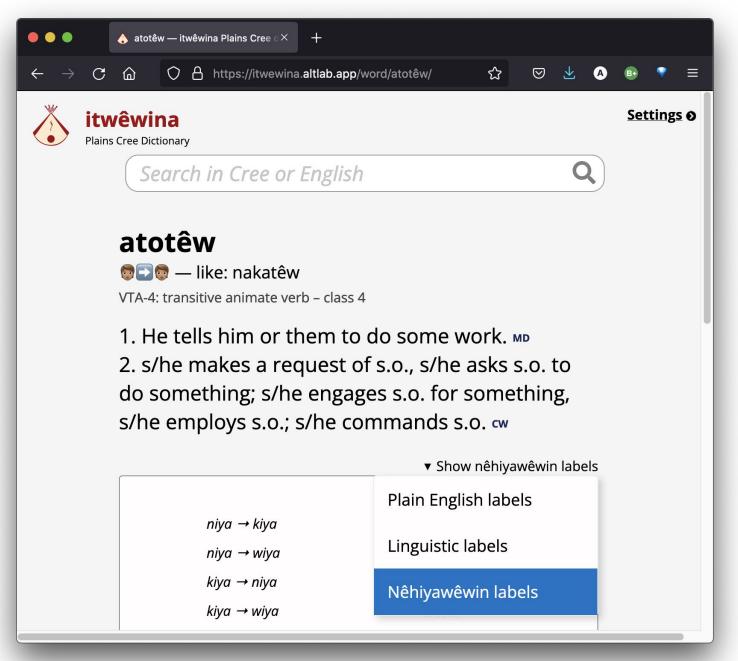
Searching with an English multi-word concept: make (someone) work



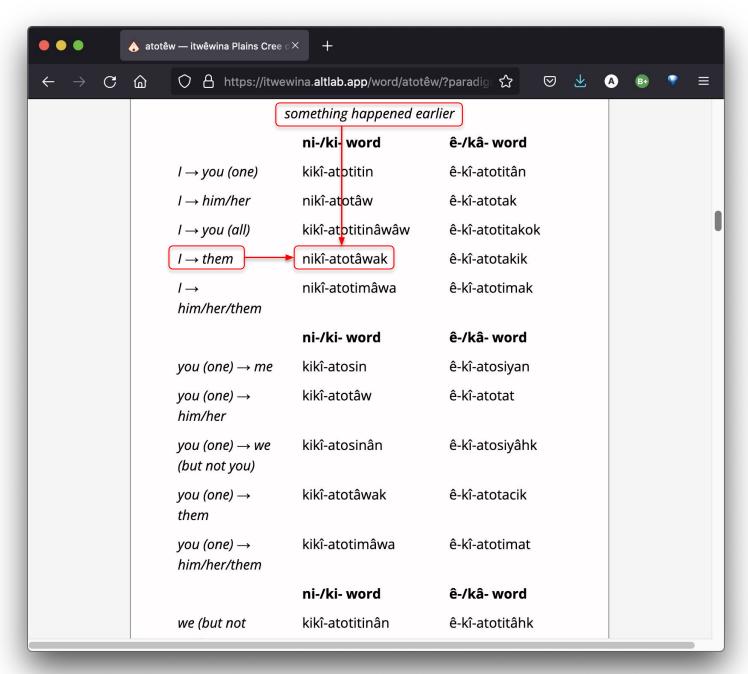
Looking up an inflected word-form in the paradigm: *I made them work*



Changing the paradigm labeling to linguistic (scientific)



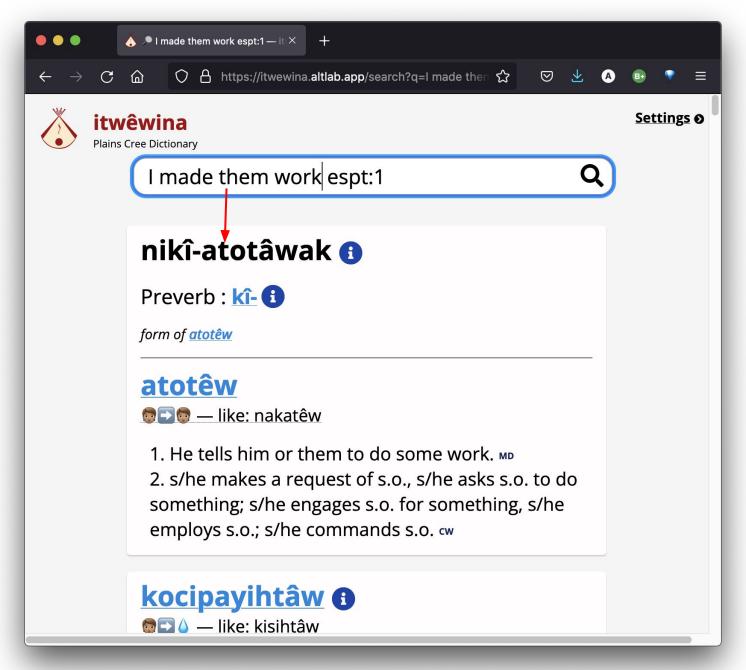
Changing the paradigm labeling to plain *nêhiyawêwin*



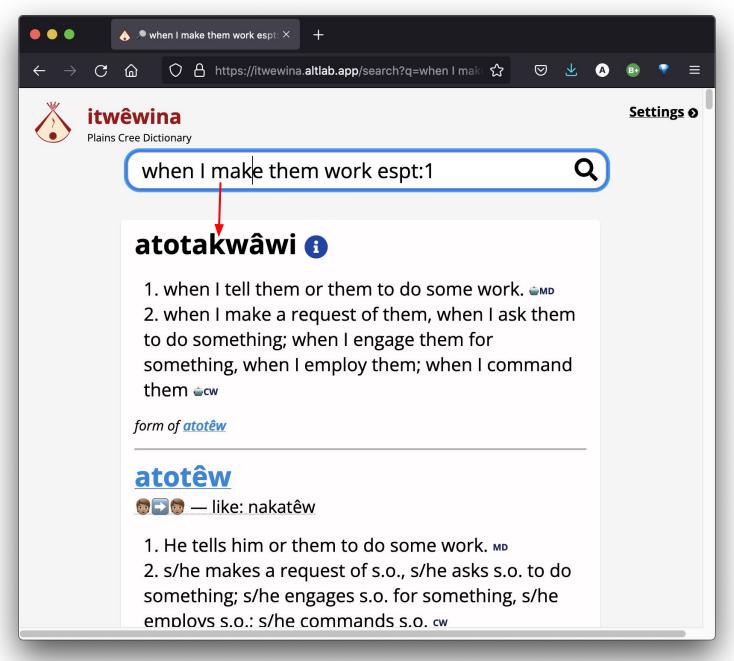
Looking up an inflected word-form in the paradigm: Past+I+them \rightarrow I made them work

How do I say X in Cree? – Part 2

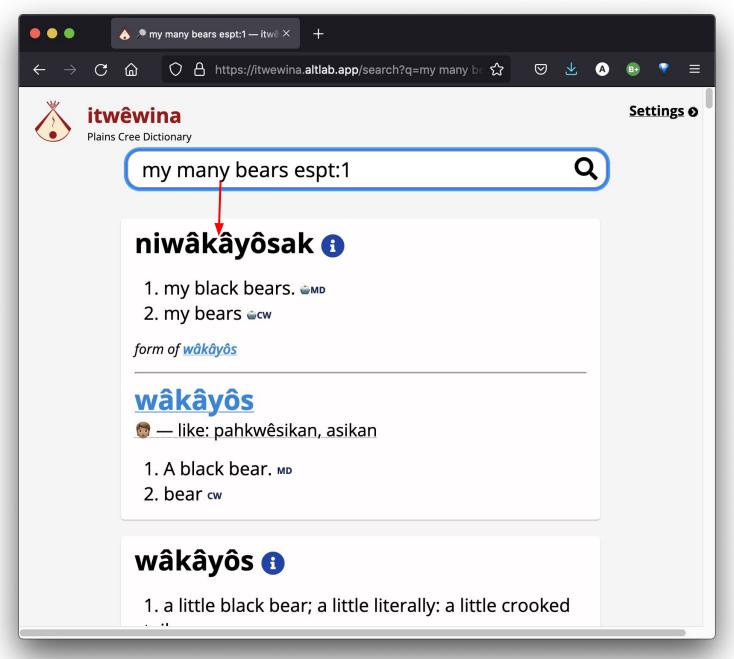
- What is the Cree word-form that corresponds to a simple English phrase?
- Is there a better solution than paradigm lookup?



Looking up with a simple English phrase: *I made them work* → *nikî-atotâwak*



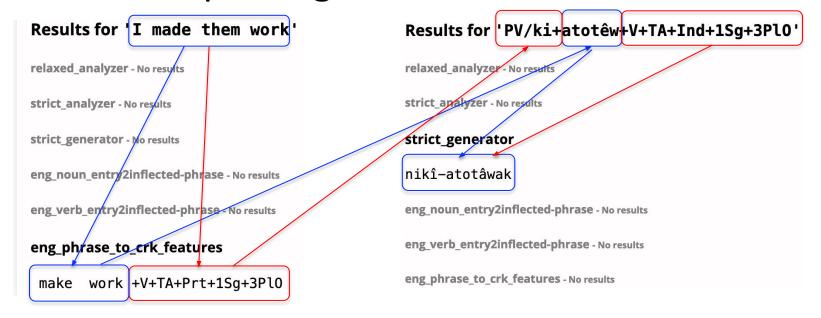
Looking up with another simple English phrase: when I make them work → atotakwâwi



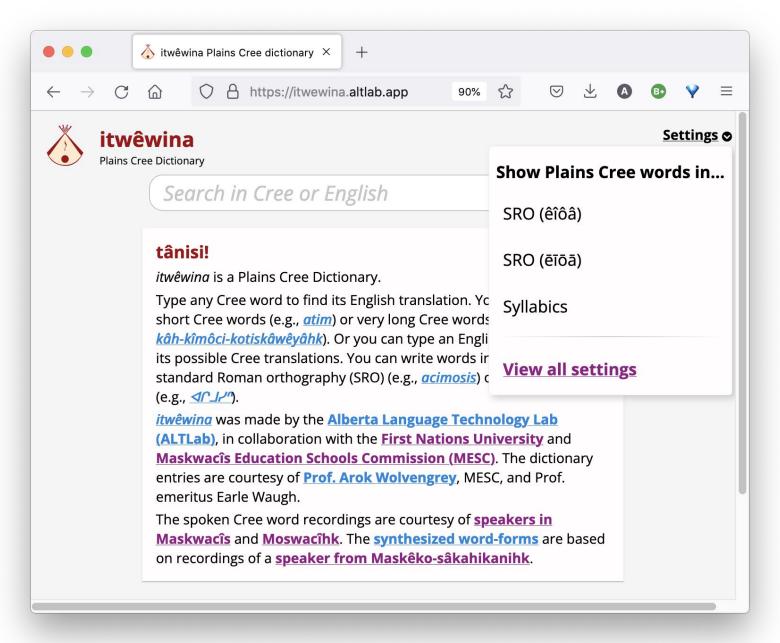
Looking up with a simple English phrase: *my many bears* → *niwâkâyôsak*

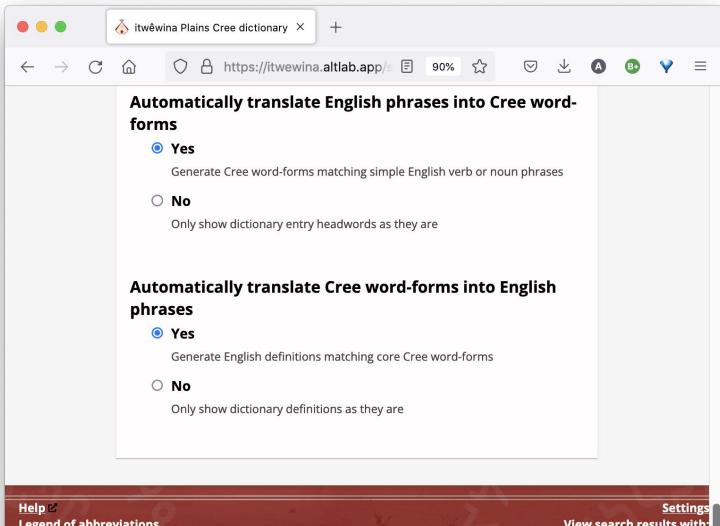
How is this achieved?

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



How is this functionality selected? → Settings





Help
Legend of abbreviations

About
Contact us

2019–2022 © Alberta Language Technology Lab.

Modified icons copyright © 2019 Font Awesome, licensed

under CC PV 4 0

Settings
View search results with:
Plain English labels
Linguistic labels

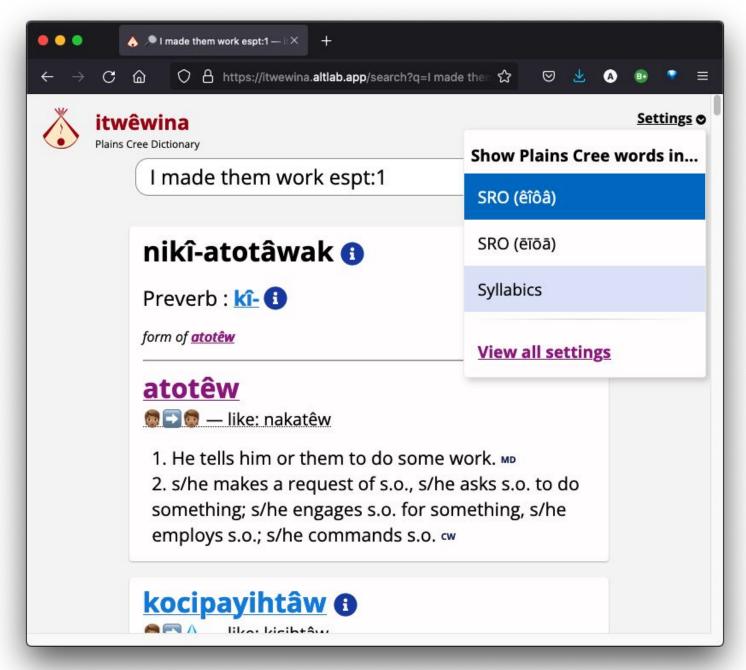
- Verb phrases (linguistic labels)
 - Subject
 - first/second/third person singular/plural/obviative, unspecified
 - Object
 - first/second/third person singular/plural/obviative, unspecified
 - Tense/aspect
 - Present, Past (kî-), Future Definite (ka-), Future Intentional (wî-), Future Conditional, Infinitive/Irrealis (ka-/ta- + Conjunct), Imperative (Immediate + Delayed)

- Verb phrases (plain English labels)
 - Actor
 - I, you, he, she, it, we, you and we, you all, they, someone
 - Goal
 - me, you, him, her, it, us, you and us, you all, them, someone
 - Time/completion
 - something happens now / someone does something now
 - s.t. happened / s.o. did s.t. earlier (kî-)
 - s.t. will certainly happen / s.o. will certainly do s.t. (ka-)
 - s.t. is going to happen / s.o. is going to do s.t. (wî-)
 - when s.t. happens / when s.o. does s.t.
 - for s.t. to happen / for s.o. to do s.t. (ka-/ta- + Conjunct),
 - let s.o. do s.t. now
 - let s.o. do s.t. later

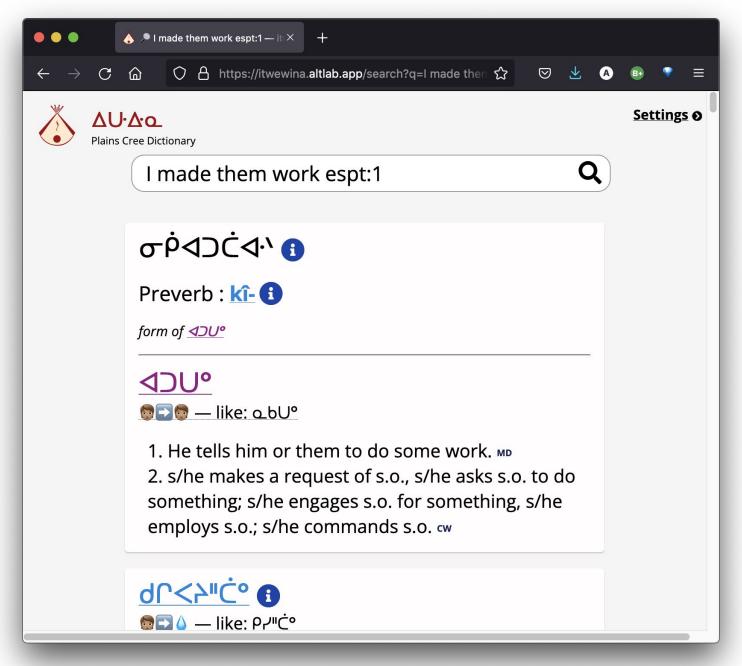
- Noun phrases (linguistic labels)
 - Non-possessed
 - singular, plural, obviative, locative
 - Possessed
 - Possessor: first/second/third person singular/plural/obviative, unspecified
 - singular, plural, obviative, locative, (distributive)

- Noun phrases
 - Non-possessed
 - one, many, further, in/on/to
 - Possessed
 - "Owner": my, your, his, her, its, our, your and our, your, their, someone's
 - one, many, further/another, in/on/to, (among)

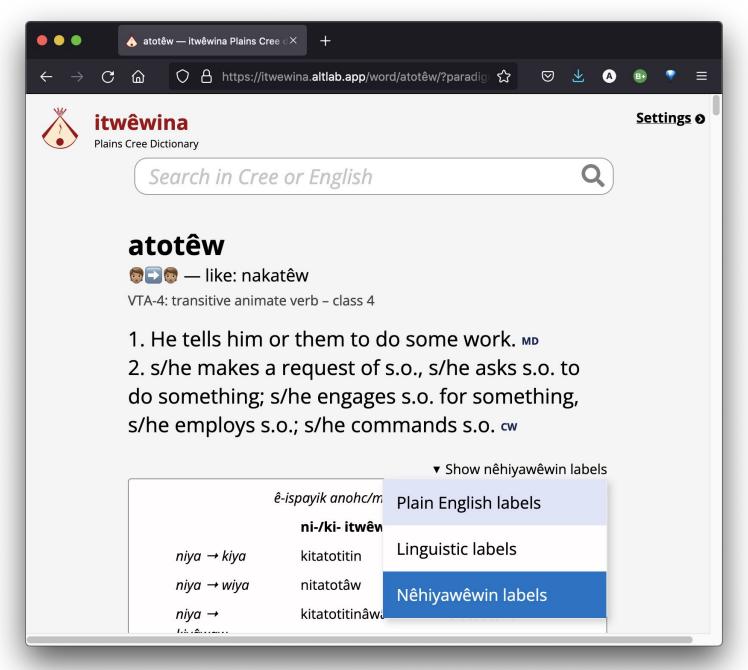
How do I write X in Cree syllabics?



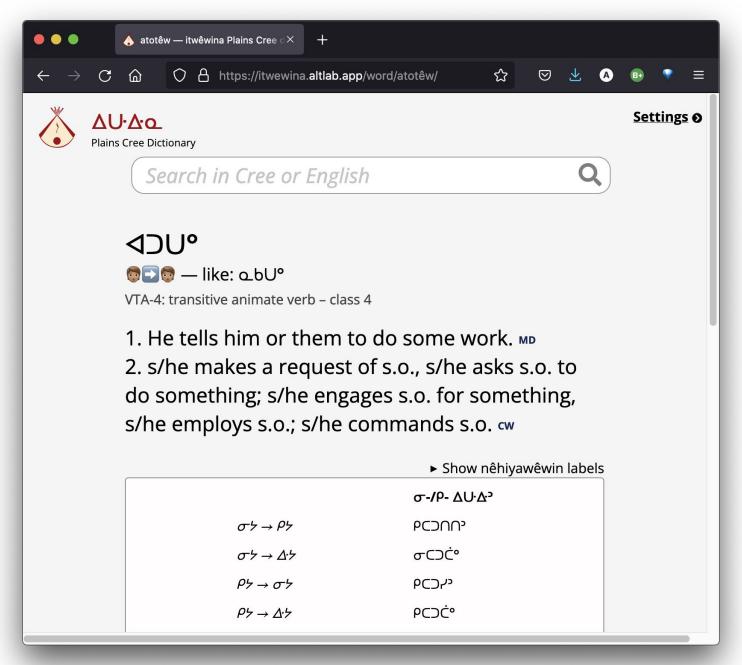
Choosing Western Cree syllabics as the display orthography for the Cree content



Choosing Western Cree syllabics as the orthography of the Cree content

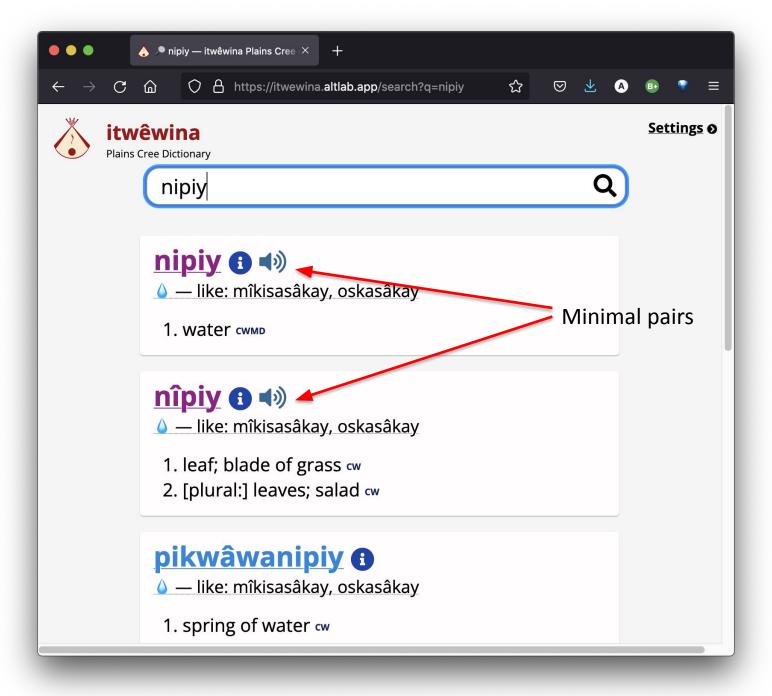


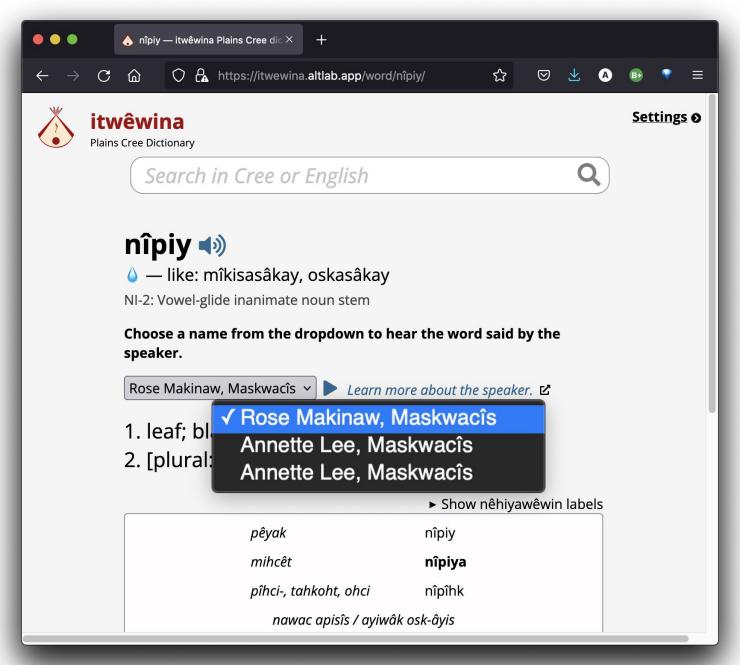
Choosing Western Cree syllabics as the display orthography for the Cree content



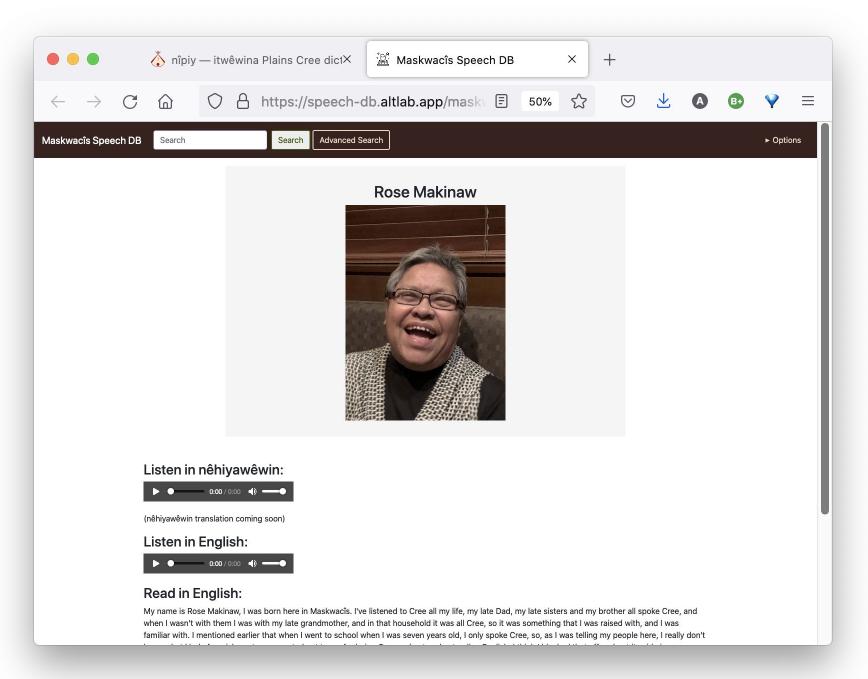
Choosing Western Cree syllabics as the orthography of the Cree content – also in the paradigms

How do I SAY X in Cree?

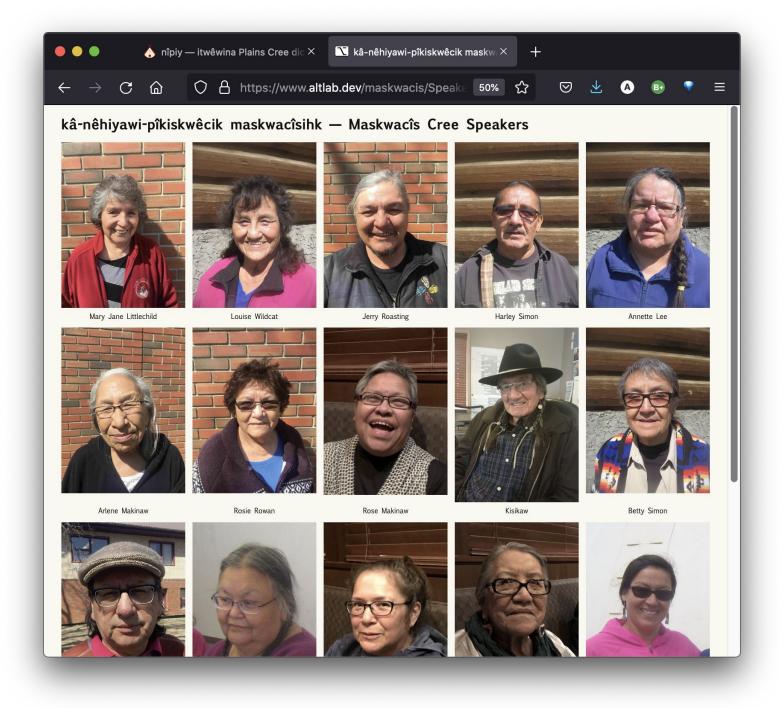




Selecting another speaker for their recording

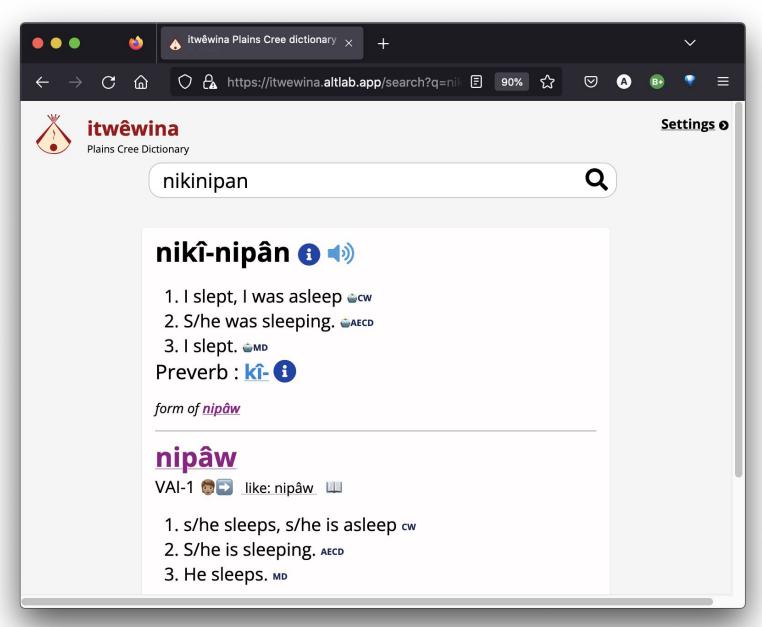


Looking up the speaker page for Rose Makinaw (Maskwacîs)

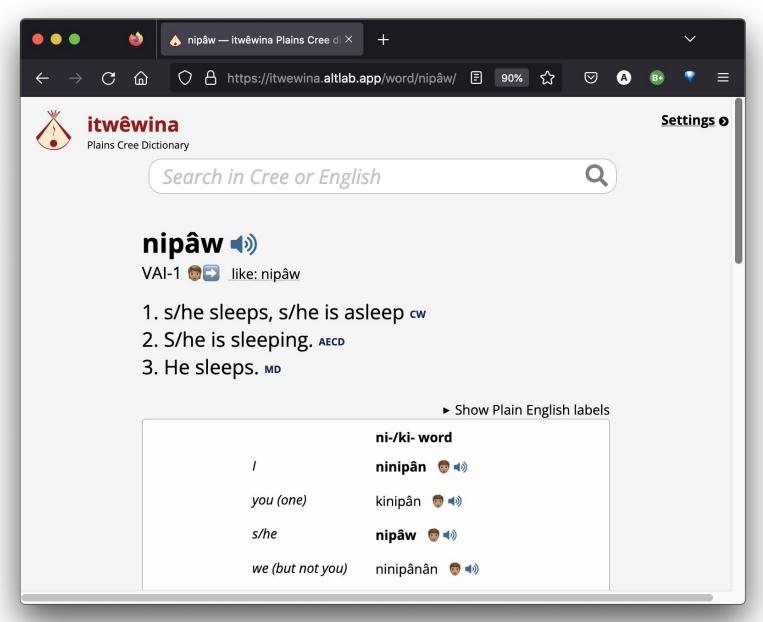


Spoken Paradigms

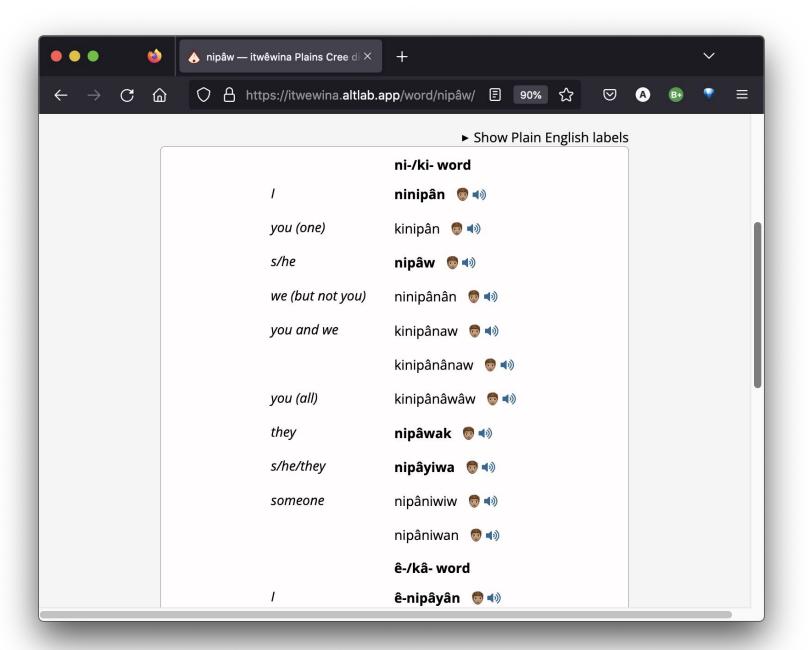
by humans and synthesizer

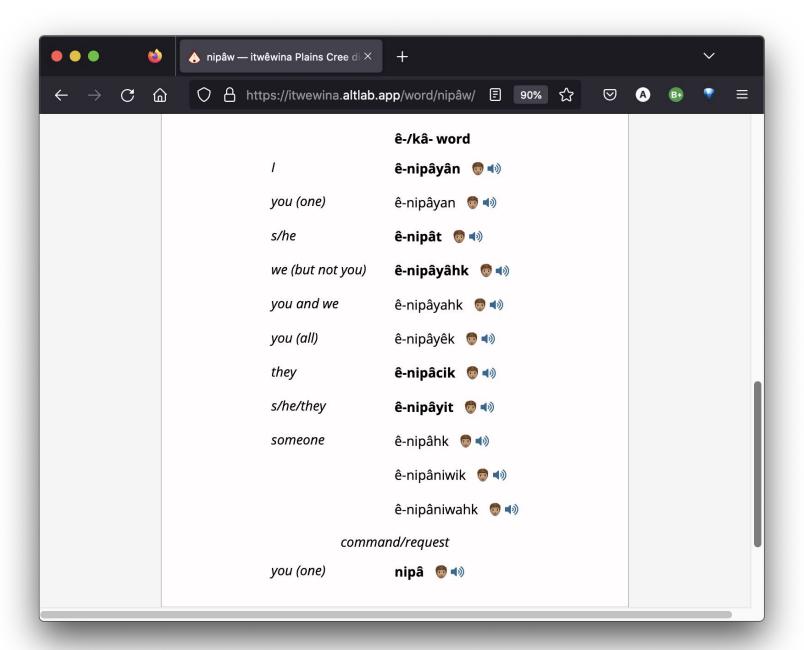


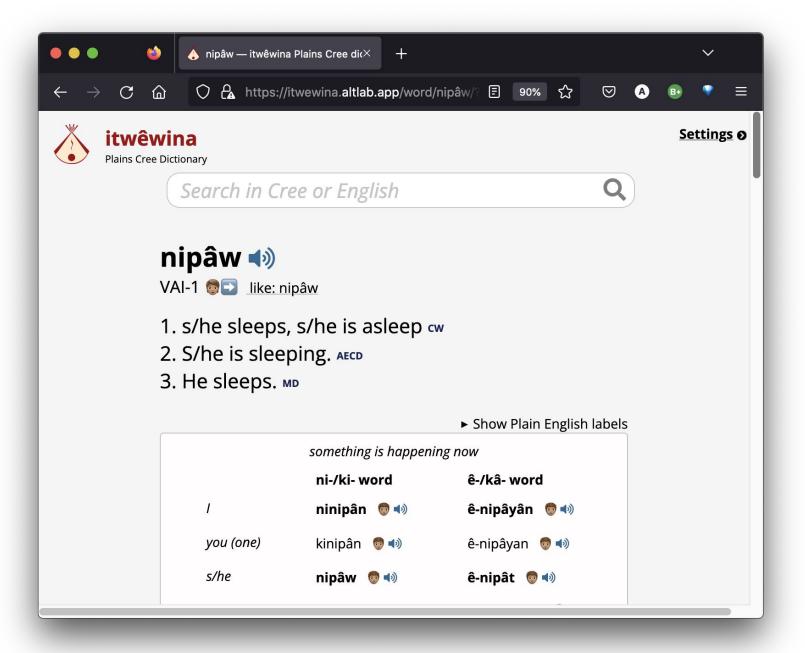
For many inflected word-forms, we do not have a human recording, but for some **selected** nouns and verbs we have full spoken paradigms, with either **human** or **synthesized** recordings.

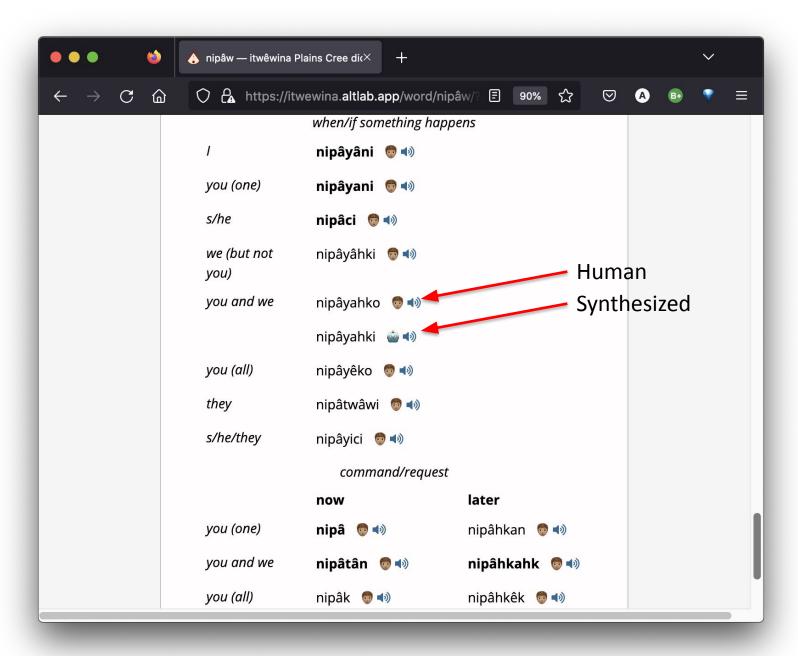


Human recordings are indicated in the paradigms with the person emoji next to the speaker emoji

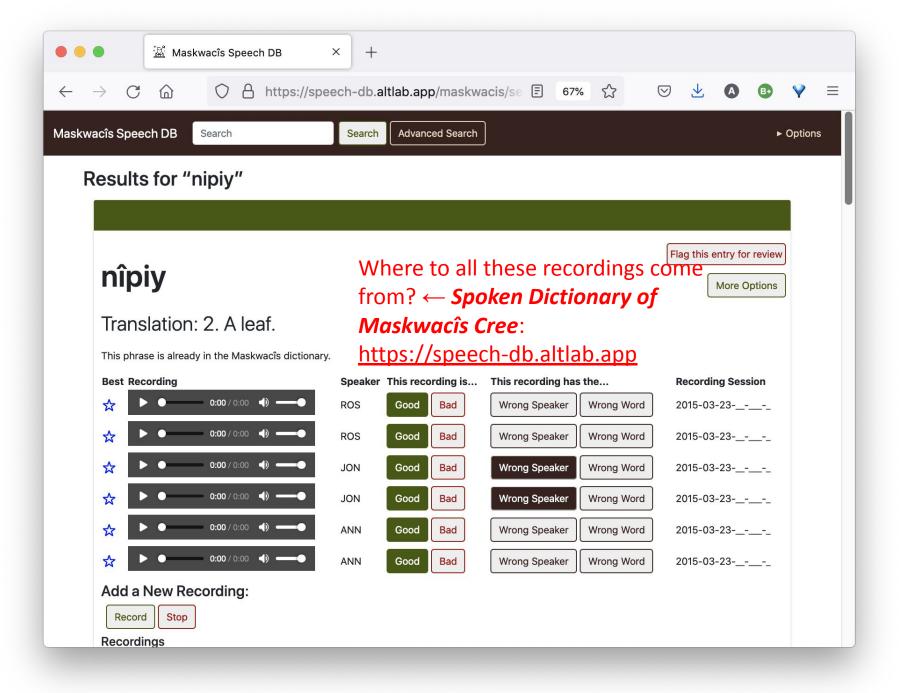






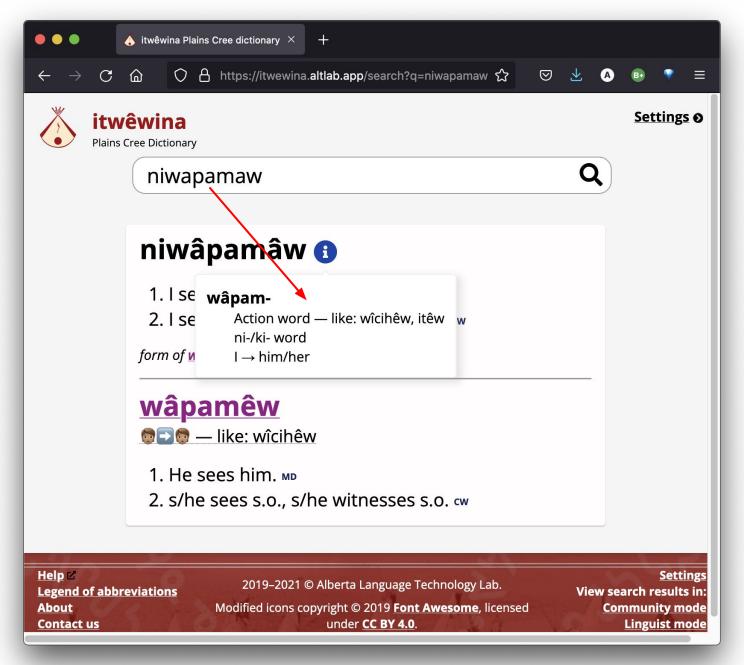


Synthesized recordings are indicated in the paradigms with the robot head emoji next to the speaker emoji

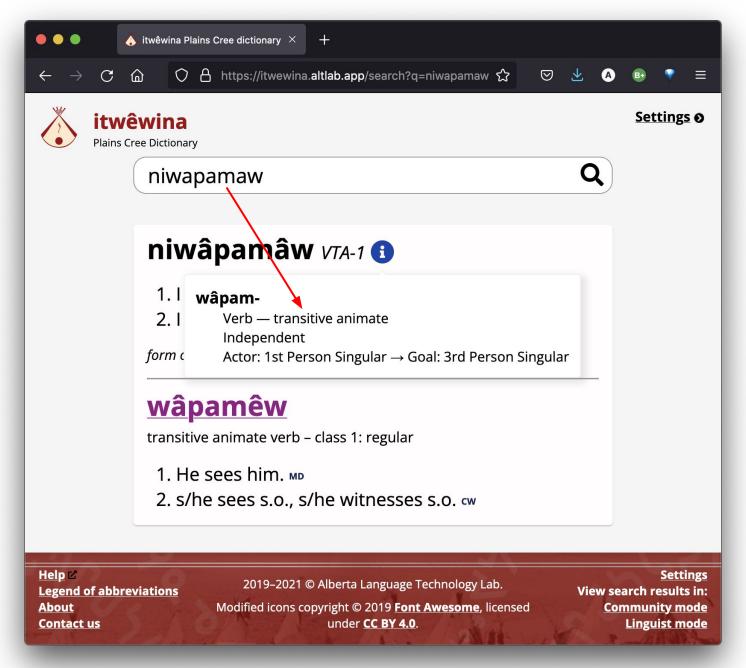


What does X (Cree) mean (in English)?

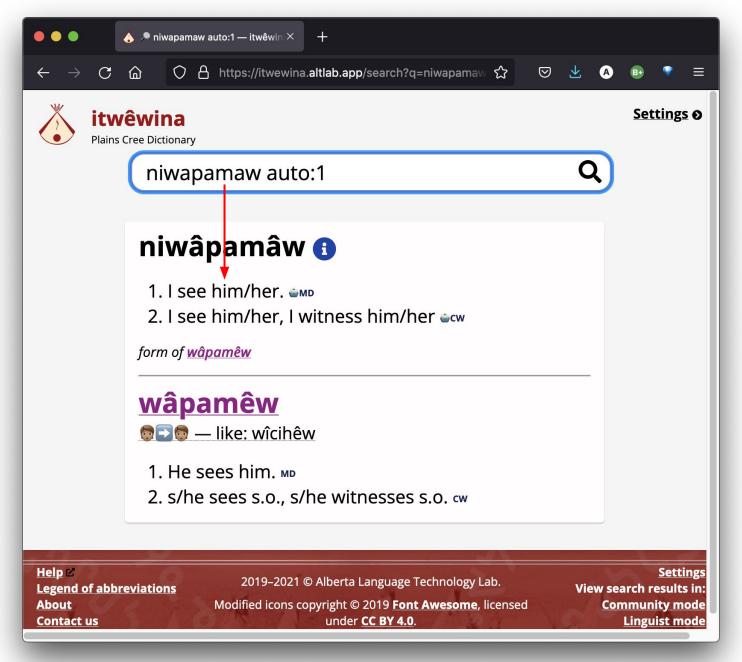
 Providing a simple English phrase as a translation, rather than a linguistic analysis



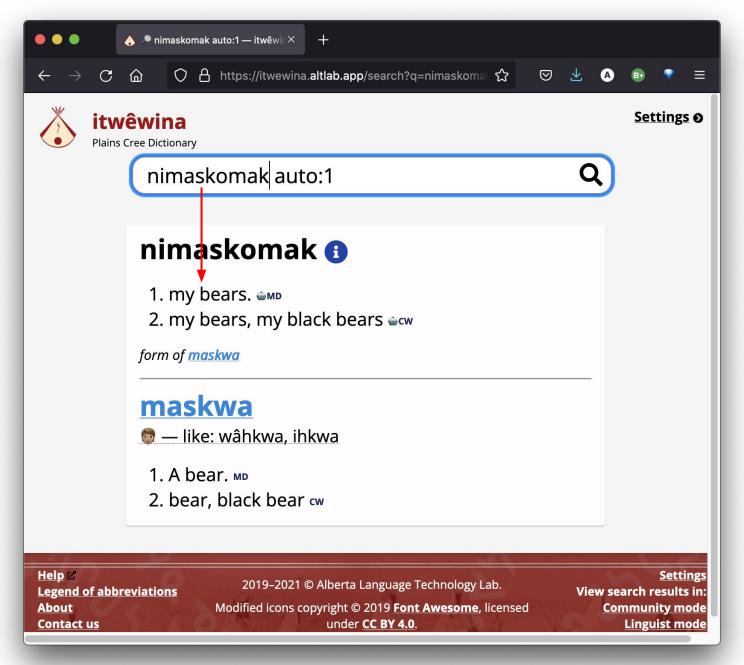
Looking up the relatively common Cree word-form: *niwâpamâw* \rightarrow *analysis in pop-up*



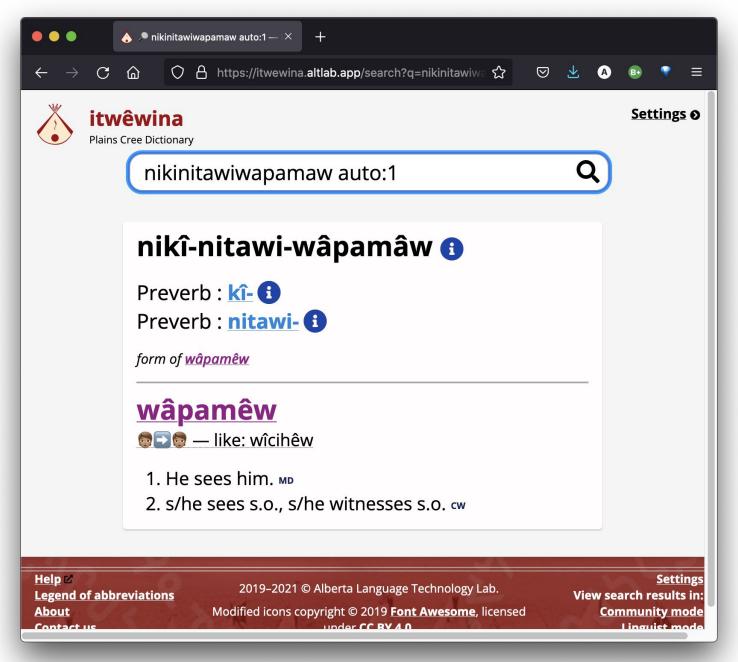
Swapping the labeling of linguistic features to: *linguistic labels*

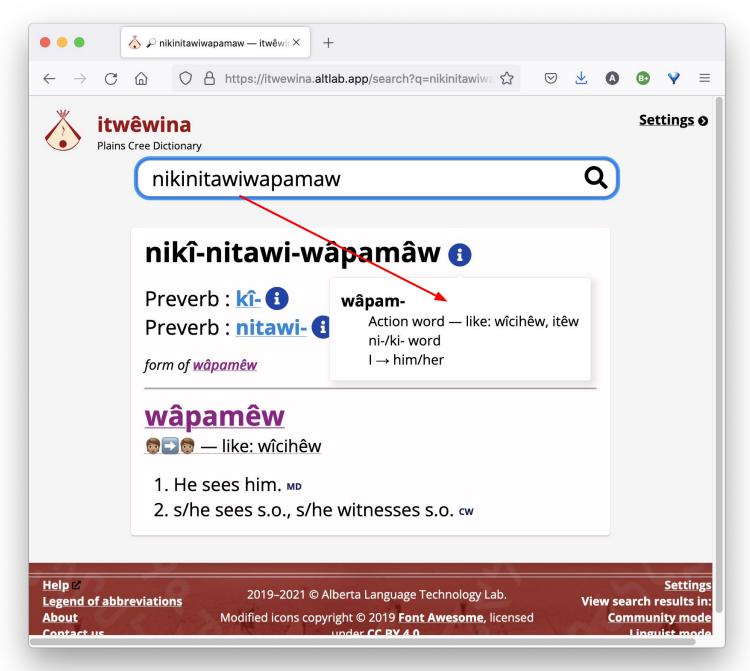


Looking up the relatively common Cree word-form: *niwâpamâw* → *I see him/her, ...*

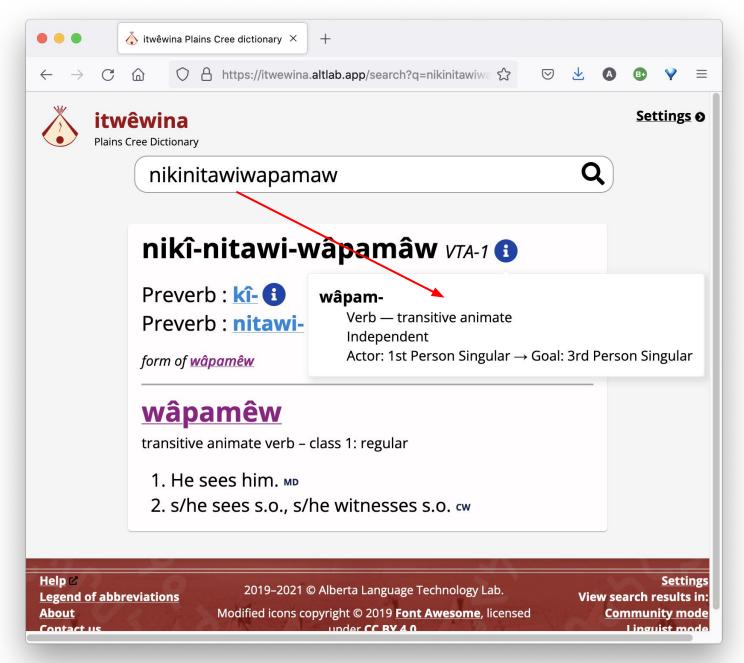


Looking up the relatively common Cree word-form: *nimaskomak* → *my bears, ...*



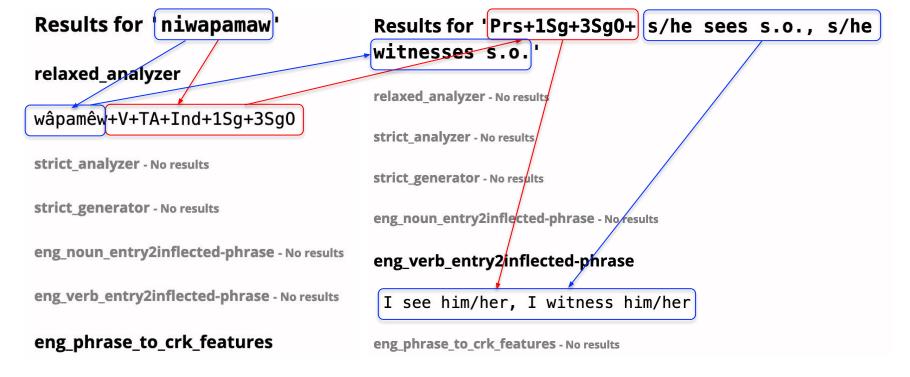


For more complex Cree word-forms: *nikî-nitawi-wâpamâw* → *only pop-up analysis available*

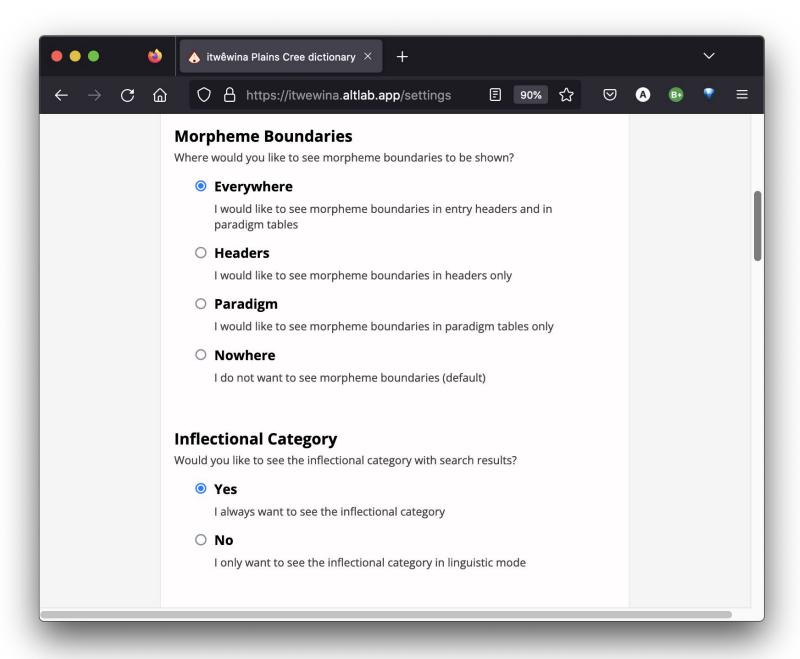


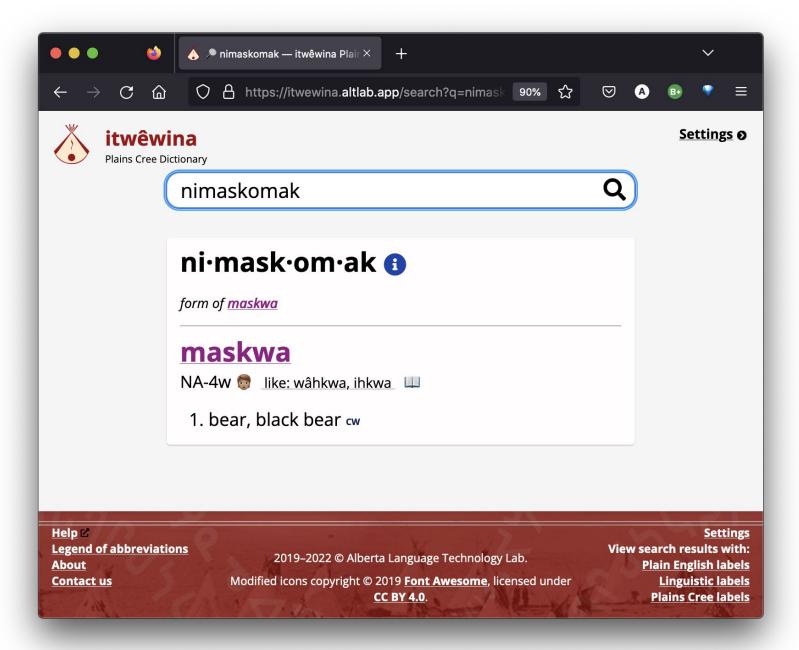
How is this achieved?

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

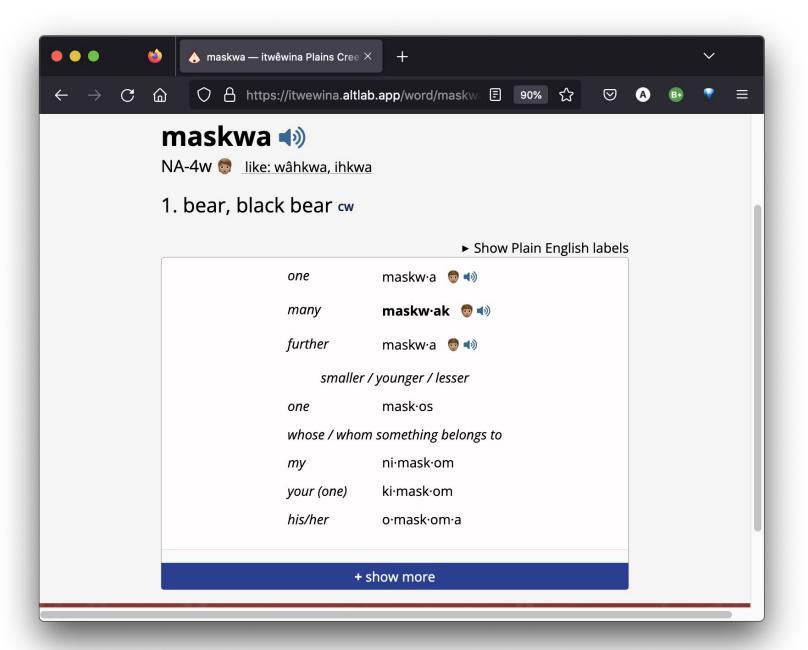


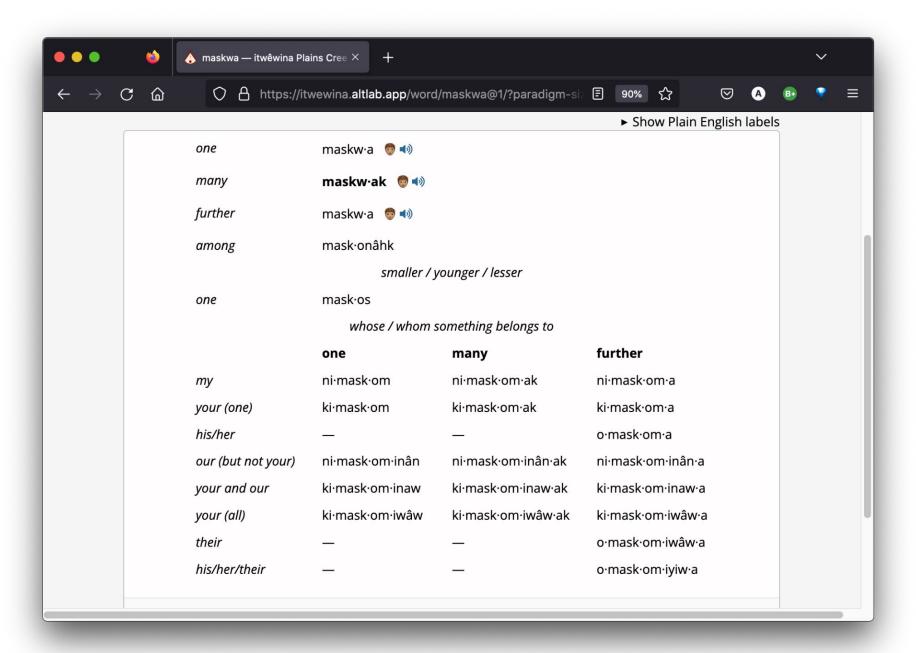
What parts are Cree words made up of?





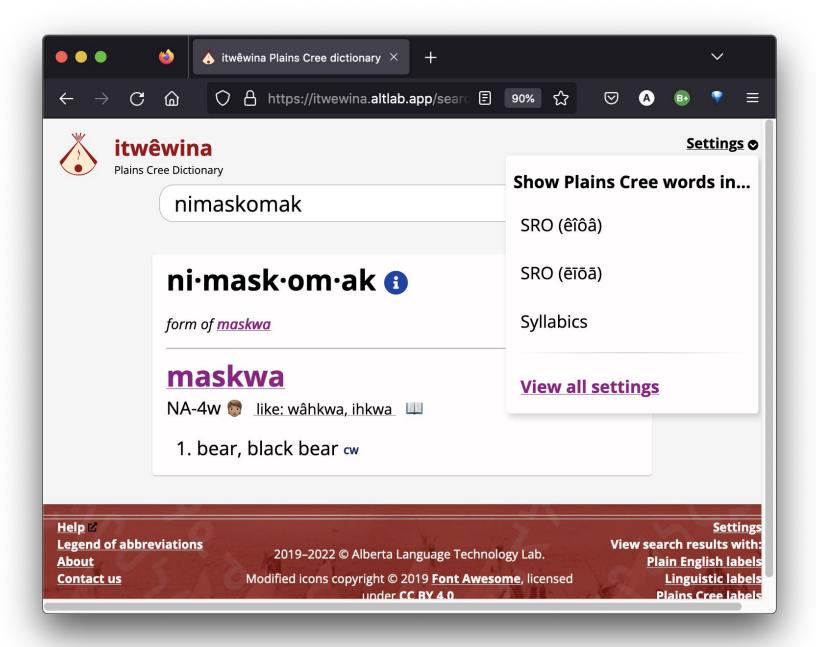
Here, the morpheme boundaries are shown for the standardized form of the search string.



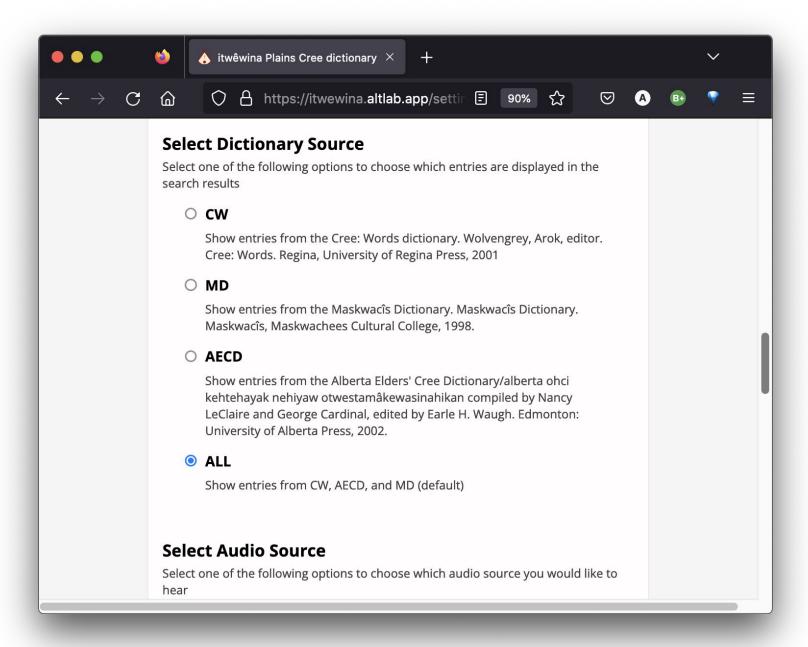


The morpheme boundaries can also be shown in all the word-forms in the paradigms.

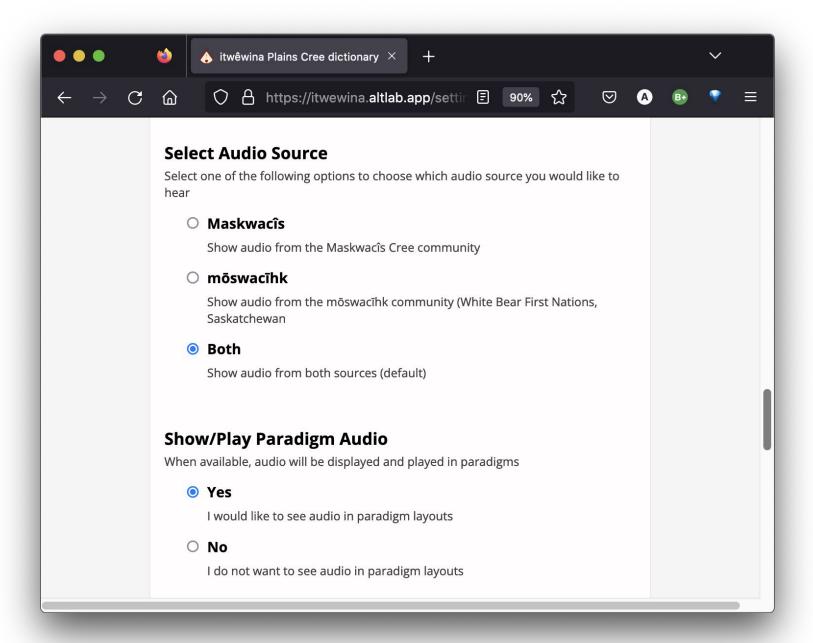
Can I choose the dictionary source or speech community?



To select the dictionary source, we go to the settings and select to view all our options.

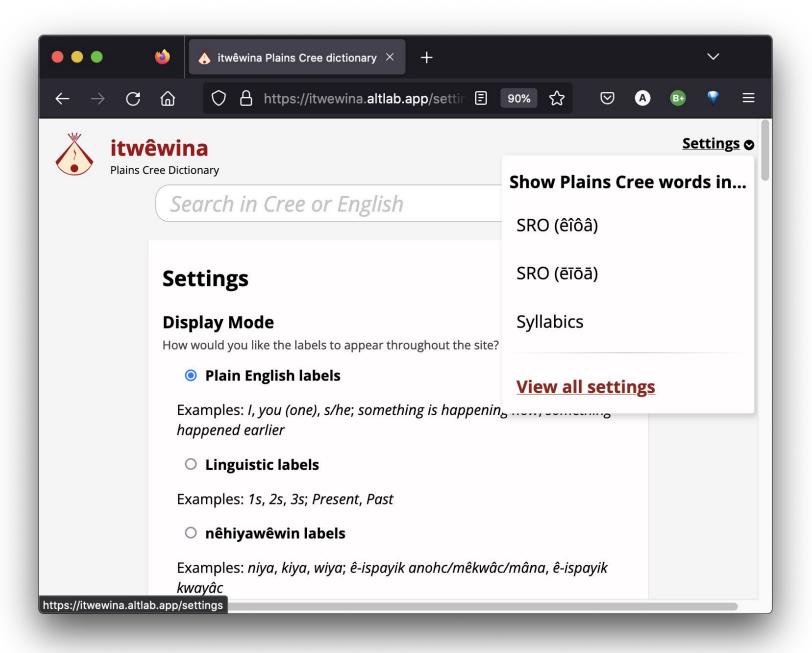


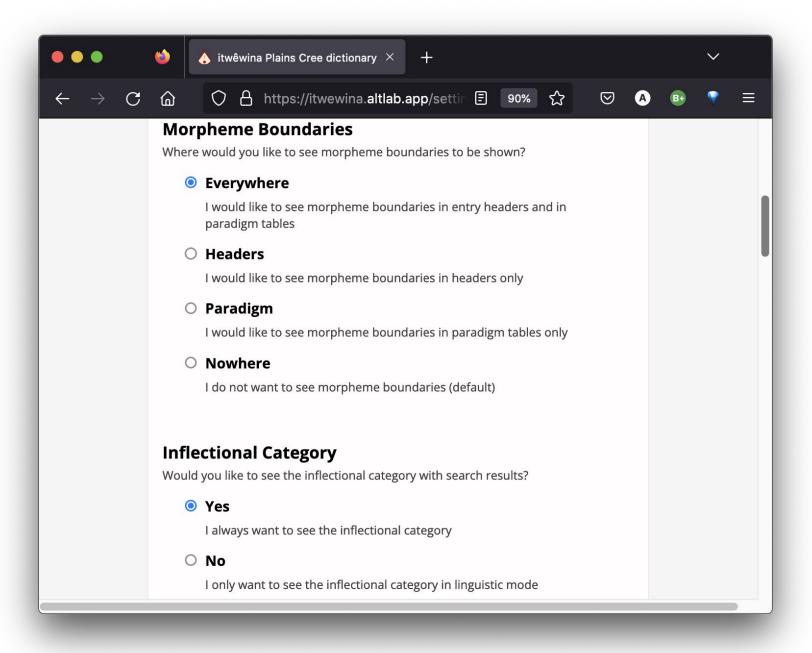
Here, we have chosen to see content from all three dictionary sources.

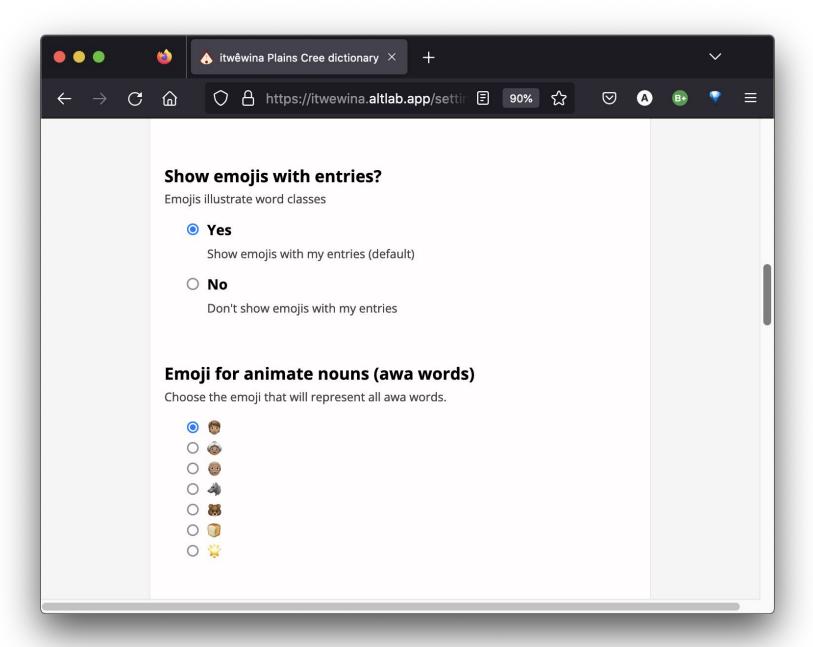


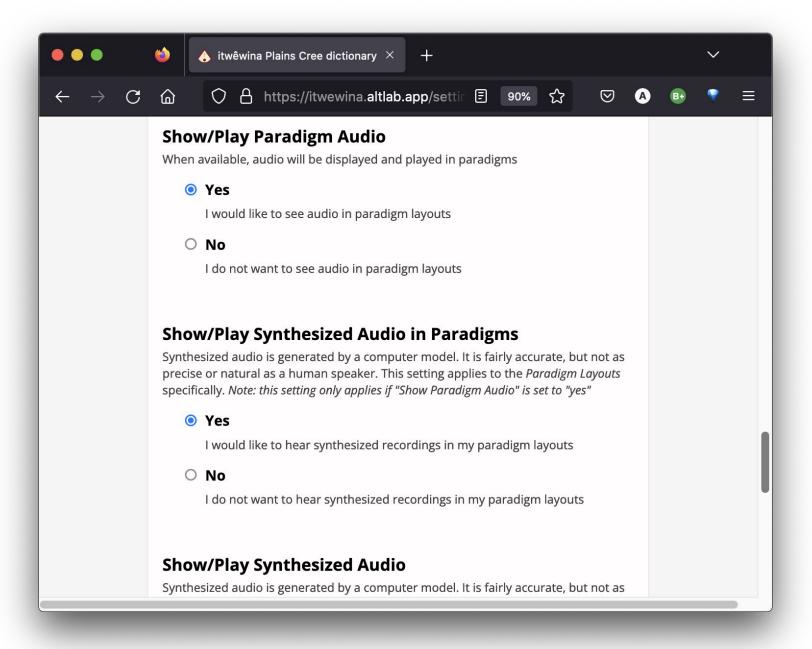
We can also choose to select which community we want to hear spoken audio from

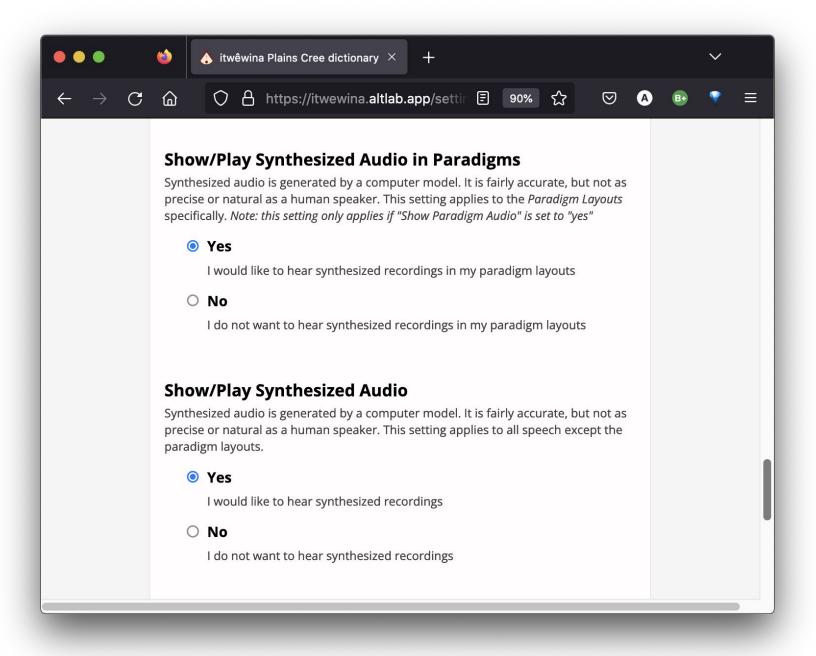
Can I choose my preferred format for how information is shown?

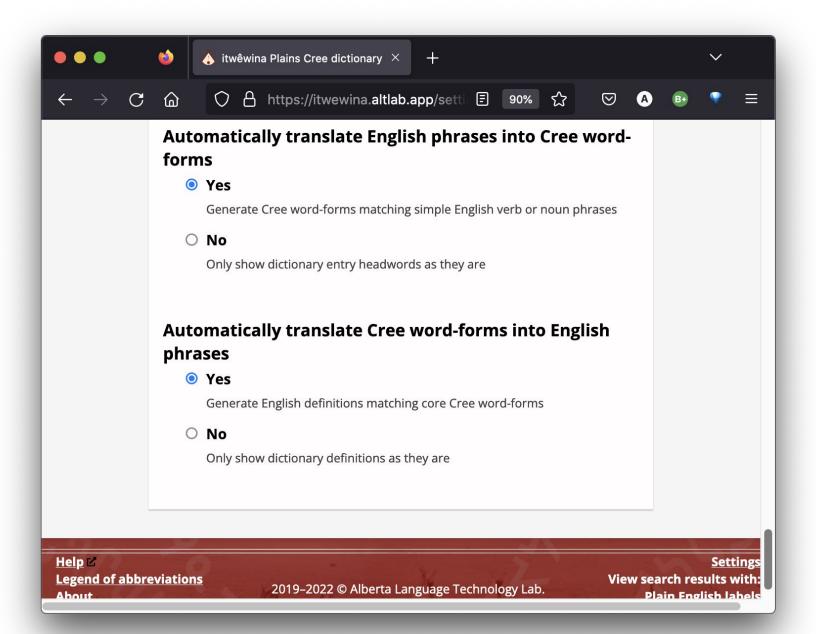


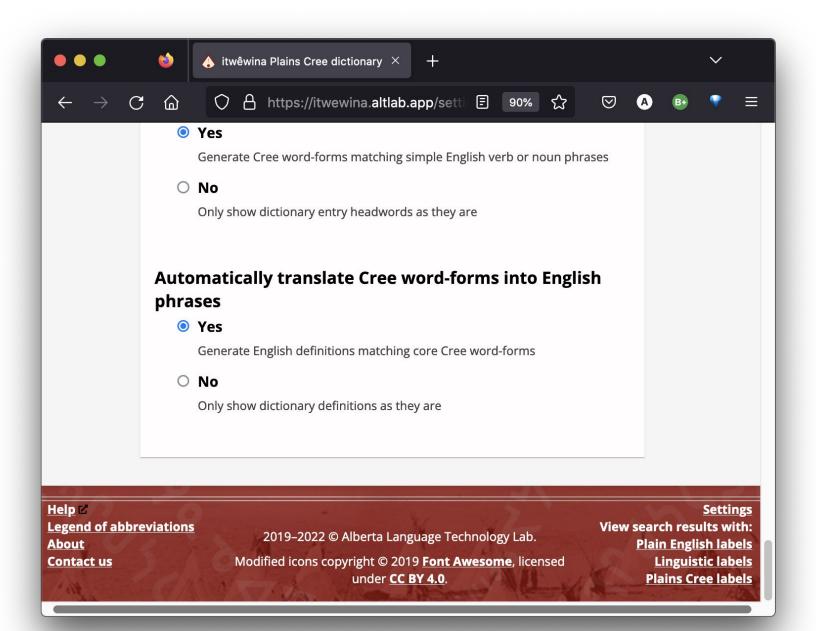












Meeting expectations – Fall 2020 \rightarrow Fall 2021 \rightarrow Fall 2021

- 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?
 - Done
 - Cree pronunciation
 - How do I say X (Cree word) in Cree?
 - Available for a part of the dictionary entries → increasing coverage of the Cree entries →
 adding fully spoken paradigms for select entries ← native speakers + synthesized
 - Cree-to-English translation
 - What does X (Cree word) mean in English?
 - good start → improving fuzzy matching

Meeting expectations – Fall 2020 \rightarrow Fall 2021 \rightarrow Fall 2022

- On-line dictionary as a language learning/instruction tool
 - Cree word building
 - What are little Cree words made up of?
 - Showing inflectional morpheme boundaries, in paradigms and throughout
 - Spoken paradigms
- Meeting the needs and expectations of multiple user communities
 - Community identity
 - user can select dictionary or speech community
 - Cree-to-English content added from Alberta Elders' Cree Dictionary
 - Fully implemented optionality for different user groups
 - user can select preferred format for any feature with multiple available options

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

- Available for testing on iOS devices via Testflight
 - https://testflight.apple.com/join/E7uvQ1i2
- Android version under development (minimally functional prototype already exists)

09:41 Tue Jan 9 ııll **奈** 80% □



Troms Reise



TFT Mobillett





Optik TV



Lyft



Uber



Conference&E..









DivvunDev...



SkyView Lite











ArriveCAN



Mehiläinen





MyUAlberta

















itwêwina offline



nikinipan

Settings o

nikî-nipân 🚯

Preverb : kî- 1

form of nipâw

<u>nipâw</u>



- 1. He sleeps. MD
- 2. s/he sleeps, s/he is asleep cw

Other possible/necessary next developments?

- On-line dictionary as a language learning/instruction tool
 - Cree word building
 - Showing derivational morpheme boundaries
 - Showing information about the individual morphemes
 - Distinguishing search results by English senses, when searching with an English word(s) → "true" English-to-Cree search
 - Organizing and splitting paradigms into smaller, more manageable panes that can be expanded and collapsed
 - Semantic search of dictionary entries (← semantic classification already exists ← WordNet / RapidWords)
 - Linguistic reference data → "grammar" pages
- Meeting the needs and expectations of multiple user communities
 - Choice of application interface language: English → Cree / nêhiyawêwin

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: https://itwewina.altlab.app/

Instructions:

https://altlab.artsrn.ualberta.ca/itwewina/

Source code:

https://github.com/UAlbertaALTLab/morphodict/