

# Finding words that aren't there

Using word embeddings to  
improve dictionary search for low-resource languages

Antti Arppe  
Andrew Neitsch  
Jolene Poulin  
Daniel W. Hieber  
Atticus G. Harrigan  
Daniel B. Dacanay

# Overview

- Background: 21C / ALTLab
- General desiderata in dictionary search
- Endangered/Indigenous languages in focus
- Dictionaries for endangered vs. endangered languages
- Vector-space search
- Qualitative + Quantitative results
- Next steps

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

SSHRC Partnership Grant 2019-2026

PD Antti Arppe

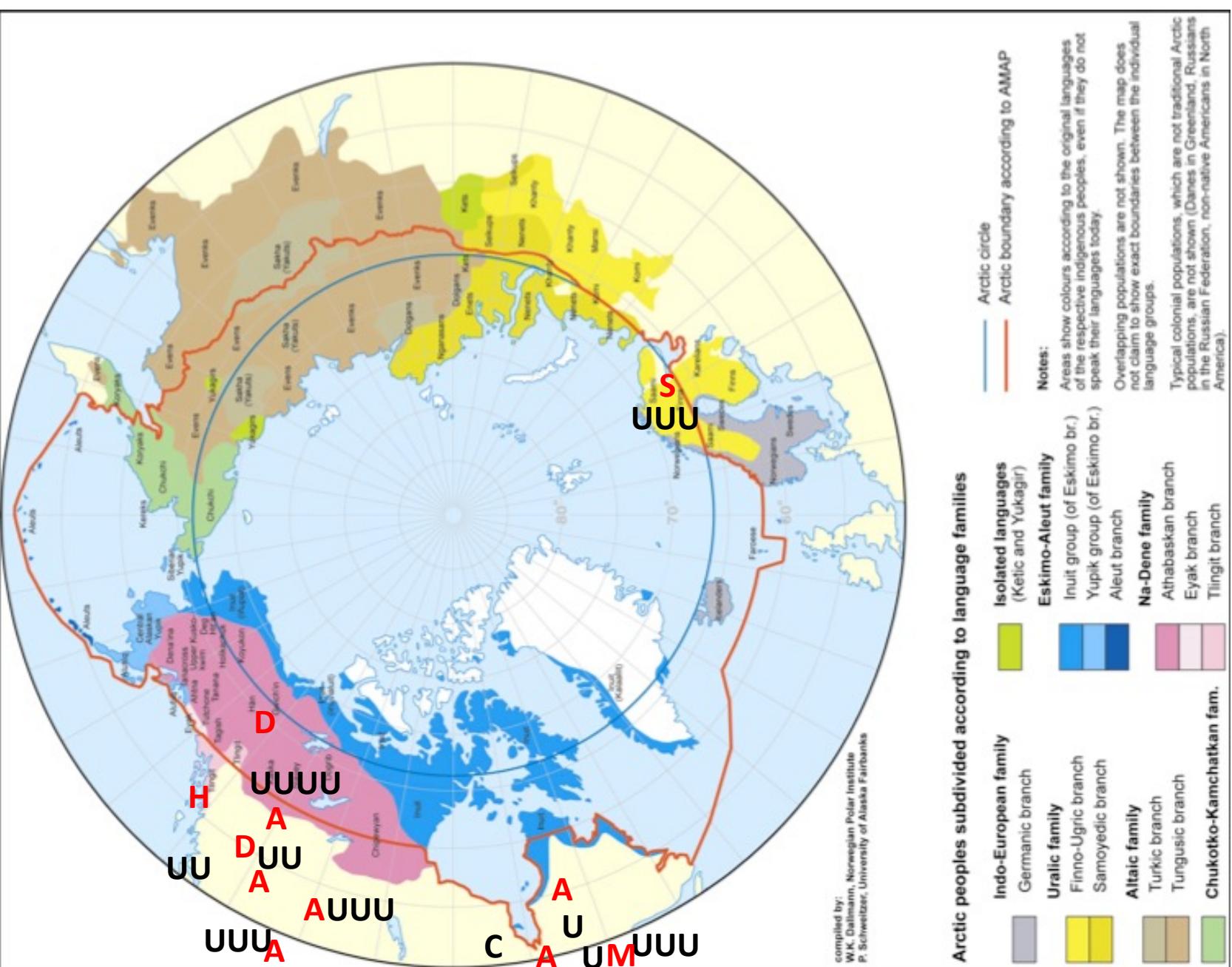
Alberta Language Technology Lab

University of Alberta

URL: <https://21c.tools>, <http://altlab.artsrn.ualberta.ca>

# Goals

- This Partnership project aims at
  - supporting **the revitalization** and **sustained daily use**
  - of **multiple Indigenous languages** in Canada
  - in **all spheres of life**
  - by **developing modern language technology**
  - in **partnership** with their **communities**
- Partnership of
  - 10 institutions
  - 27 individuals



University + Community Partners + Participants

# Partnership: Participants – Languages / Communities

- Uralic
  - **Saami languages** → UiT – Arctic University of Norway
- Algonquian
  - **Plains Cree** ([Maskwacîs Education Schools Commission](#)) ↗ UAlberta, First Nations University, Cree Literacy Network
  - **Woods Cree** → FNUC; UAlberta → [Lac La Ronge Indian Band](#)
  - **Blackfoot** → Lethbridge
  - **Arapaho** → Boulder (Colorado, US)
  - Algonquian Dictionaries Infrastructure: **Odawa, Northern East Cree** → Carleton
- Dene
  - **Tsuut'ina** ([Tsuut'ina Gunáhà Institute, Tsuut'ina Language Commissioner](#)) → Carleton
  - **Upper Tanana** → USaskatchewan
  - **Dene Suliné** UAlberta / USaskatchewan ([Tlicho Yatíi](#))
- Other
  - **Haida** → UAlberta
  - **Nakoda (Siouan)** → UAlberta
  - **Mohawk (Iroquoian)** → National Research Council

# Partnership: Participants – Universities

- Alberta ([Arppe, Lachler, Thunder](#))
  - [Plains Cree](#) ([Maskwacîs Education Schools Commission, MESC](#): Eileen Marthiensen, Rose Makinaw)
  - [Haida](#) ([Haida Guu.suu Society](#): Lucy Bell)
  - Community-instructor training (expanding CILLDI)
- Carleton ([Junker, Cox](#))
  - [Algonquian Dictionaries Infrastructure](#) ↗ extended functionalities through updated technology
  - [Tsuut’ina](#) ([Tsuut’ina Gunáhà Institute, Office of the Tsuut’ina Language Commissioner](#): Bruce Starlight)
- Colorado, Boulder ([Hulden, Cowell, Palmer](#))
  - Computational Workshops for Indigenous languages (CWILs)
  - [Arapaho \(Algonquian\)](#), Siouan
- First Nations University of Canada ([Wolvengrey, Ratt](#)) & University of Saskatchewan ([Lovick](#))
  - Extensive documentation work on [Plains Cree](#), [Woods Cree](#), [Upper Tanana](#)
- UiT – Arctic University of Norway (Trosterud, Moshagen, Antonsen)
  - [Saami languages](#) (Giellatekno, Divvun)
  - Original Giella development infrastructure
    - Language-independent application solutions (on-line dictionaries, spell-checkers, grammar-checkers)
- Wisconsin, Madison (Valentine) & Sudbury (Corbiere)
  - [Ojibwe/Odawa](#) ([Odawa communities](#): Mary Ann Corbiere, Alan Corbiere)
- Cree Literacy Network (Ogg)
- National Research Council (Kazantseva, Kuhn)
  - Support for Canadian infrastructure development
  - [Mohawk \(Iroquoian\)](#) school-board connections

# Software resources and applications – “low-hanging fruit”

- (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



## Useful (on-line) dictionary:

Finding results for all types of words and phrases,  
including fancier and/or rarer or newer words

E.g.

*collaborate, procrastinate, cellphone, blackberry,  
nokia*

Home : Oxford English Dictionary X +

ABOUT COMMUNITY BLOG

OED | Oxford English Dictionary  
*The definitive record of the English language*

Subscribe

Discover the story of English  
More than 600,000 words, over a thousand years

More about the OED »  
Print edition »

Welcome to OED Online. If you or your library subscribes, dive straight in to the riches of the English language. If not, click on the images below to learn more about the OED, see **What's new**, or take a look at **Aspects of English**, our language feature section.

Why do we call a moment of sudden realization or illumination a 'lightbulb'?



collaborate, v. : Oxford English Dictionary

ABOUT COMMUNITY BLOG Access: University of Alberta

# OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

View full results for 'collaborate' Help on Dictionary Entry | Print | Save | Email | Cite

## collaborate, *v.*

View as: Outline | [Full entry](#)

Text size: A A

Pronunciation: <sup>?</sup> Brit. /kə'labərət/, U.S. /kə'læbə'reɪt/

Frequency (in current use):

Etymology: < modern French *collaborer*, or its original, Latin *collabōrāre* (< *col-* together + *labōrāre* to work, LABOUR *v.*): see *-ATE suffix<sup>3</sup>*. [\(Show Less\)](#)

**1. intransitive.** To work in conjunction with another or others, to co-operate; esp. in a literary or artistic production, or the like.

Thesaurus »

This entry has not yet been fully updated (first published 1891; most recently modified version published online March 2022).

Entry history

Entry profile

Previous version:  
OED2 (1989)

In this entry:

procrastinate, v. : Oxford English X +

View full results for 'procrastinate' Help on Dictionary Entry | Print | Save | Email | Cite

# procrastinate, v.

Text size: A A

View as: Outline | [Full entry](#)

Quotations: Show all | [Hide all](#) Keywords: On | [Off](#)

**Pronunciation:** <sup>?</sup> Brit. /prə(v)ˈkræstɪneɪt/, U.S. /prəˈkræstə,neɪt/, /prəʊˈkræstə,neɪt/

**Forms:** 1500s **procrastinate** (past participle), 1500s– **procrastinate**.

**Frequency (in current use):**

**Origin:** A borrowing from Latin. **Etymons:** Latin *prōcrāstināt-*, *prōcrāstināre*.

**Etymology:** < classical Latin *prōcrāstināt-*, past participle stem (see [-ATE suffix<sup>3</sup>](#)) of *prōcrāstināre* to put off till the next day, to defer, delay < *prō-* PRO- prefix<sup>1</sup> + *crāstinus* belonging to tomorrow (see [CRASTIN n.](#)). Compare Middle French *procrastiner*

[PROCRASTINE v.](#), Italian *procrastinare* (c1300). Compare [PROCRASTINATION n.](#), [PROCRASTINE v.](#) ([Show Less](#))

**1. intransitive.** To defer action, delay; to be dilatory.

Often with the sense of deferring through indecision, when early action would have been preferable.

Thesaurus »

This entry has been updated (OED Third Edition, June 2007; most recently modified version published online March 2022).

Entry history

Entry profile

Previous version: OED2 (1989)

In this entry:

pro'crastinated

cell phone, n. : Oxford English D X +

View full results for 'cellphone' Help on Dictionary Entry | Print | Save | Email | Cite

# cell phone, n.

Text size: A A

View as: Outline | [Full entry](#)

Quotations: Show all | Hide all    Keywords: On | Off

**Pronunciation:** <sup>?</sup> Brit. /'sel fəʊn/, U.S. /'sel foun/

**Forms:** 1900s– **cell phone**, 1900s– **cellphone**.

**Frequency (in current use):**

**Origin:** Formed within English, by compounding. **Etymons:** [CELLULAR adj.](#), [PHONE n.<sup>2</sup>](#)

**Etymology:** < *cell-* (in [CELLULAR adj.](#)) + [PHONE n.<sup>2</sup>](#)

Compare [CELLNET n.](#) and later [CELL n.<sup>3</sup>](#)

Chiefly North American.

A portable wireless telephone that transmits and receives signals via a cellular ([CELLULAR adj. 6](#)) network; a mobile phone; *esp.* (in later use) a smartphone. Cf. [CELL n.<sup>3</sup>](#) 1.

Thesaurus »  
Categories »

This entry has been updated (OED Third Edition, September 2017; most recently modified version published online December 2021).

Entry history  
Entry profile

In this entry:

- cell phone bill
- cell phone call
- cell phone charger
- cell phone company
- cell phone network
- cell phone number
- cell phone reception

Quick search results : Oxford En X +

ABOUT COMMUNITY BLOG Access: University of Albe

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

# Quick search results

Showing 1-2 of 2 results in 2 entries

**Widen search?** Find 'blackberry' in: » phrases (9) » definitions (39) » etymologies (10) » quotations (240) » full text (173)

**View as:** List | [Timeline](#)

**Sort by:** Entry | [Frequency](#) | [Date](#)

1. [blackberry, n.](#) [View full entry](#)

OE

...The edible berry-like fruit of the bramble, *Rubus fruticosus*, and its cultivated varieties (see sense 1b), which is an aggregate fruit consisting of a cluster of soft, sweet,...

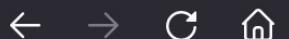
Your current search (entries):

**blackberry**

[Save search](#)  
[Refine search](#)

Refine your search

▶ [Subject](#)  
▶ [Language of Origin](#)  
▶ [Region](#)



3. Usually in form **BlackBerry**. A proprietary name for: a type of pager or smartphone capable of sending and receiving email messages.

- 1999 *Daily News* (N.Y.) 19 Jan. 20/2 Canadian tech company Research In Motion hopes to send an alert to the industry with a snazzy new pager called the BlackBerry... It allows users to send and receive E-mail, text messages and faxes, and send voice-mail.
- 2000 *N.Y. Times* 9 Nov. B9/1 [She] sent the vice president a message at 1:20 a.m. Central time on his BlackBerry.
- 2004 *Daily Tel.* 8 Dec. 6/1 If you thought that mobile telephones had made it impossible to hide from the boss,..just wait until you have tried a BlackBerry.
- 2009 *Atlantic Monthly* Apr. 81/1 I put my BlackBerry in my cargo pocket, not trusting it to the mild chaos of the cloak stand.

([Hide quotations](#))

---

## PHRASES

No Search Results : Oxford Engl X +

ABOUT COMMUNITY BLOG Access: University of Albe

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

## Quick search results

No dictionary entries found for 'nokia'.

Did you mean:

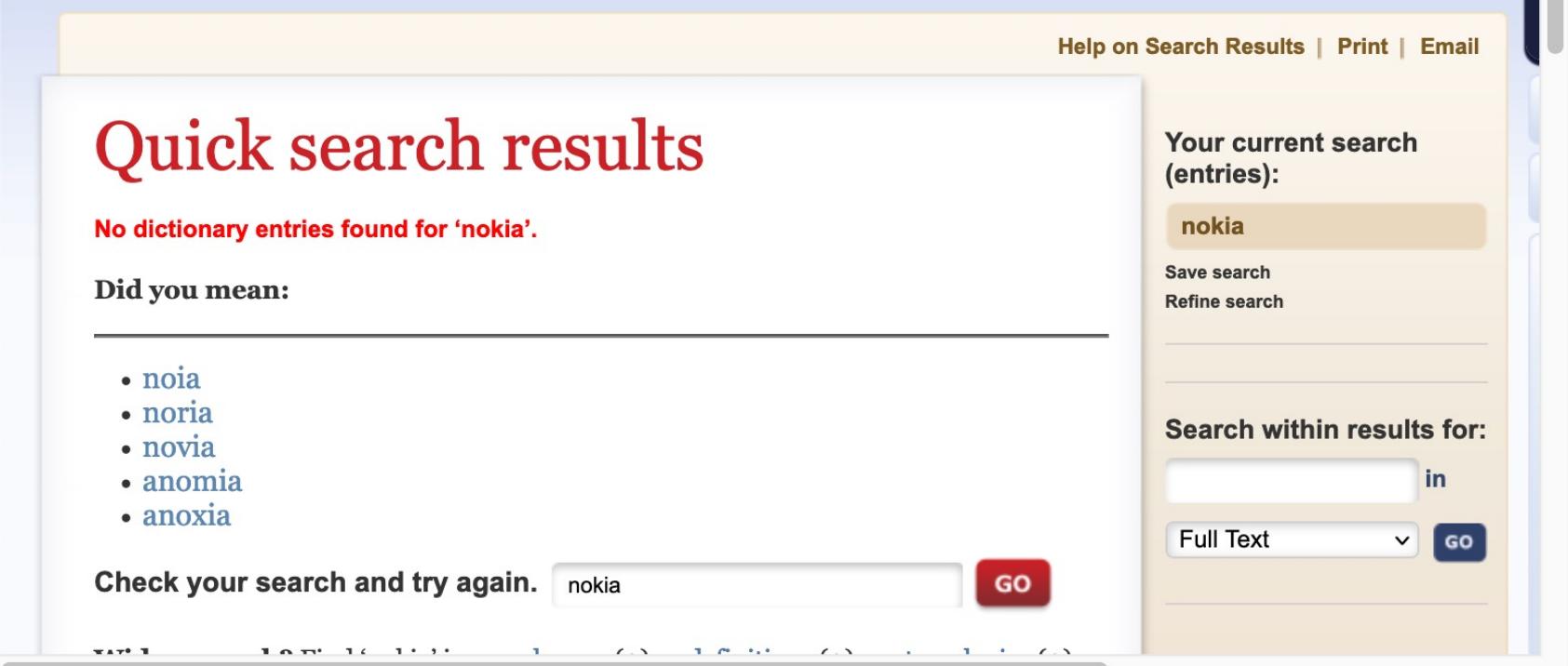
- noia
- noria
- novia
- anomia
- anoxia

Check your search and try again.  GO

Your current search (entries):  
nokia

Save search  
Refine search

Search within results for:  
 in  
Full Text GO



Useful (on-line) dictionary:

Finding results for compound phrases

E.g.

*work together*

No Search Results : Oxford Engl X +

ABOUT COMMUNITY BLOG Access: University of Albe

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

## Quick search results

No dictionary entries found for 'work together'.

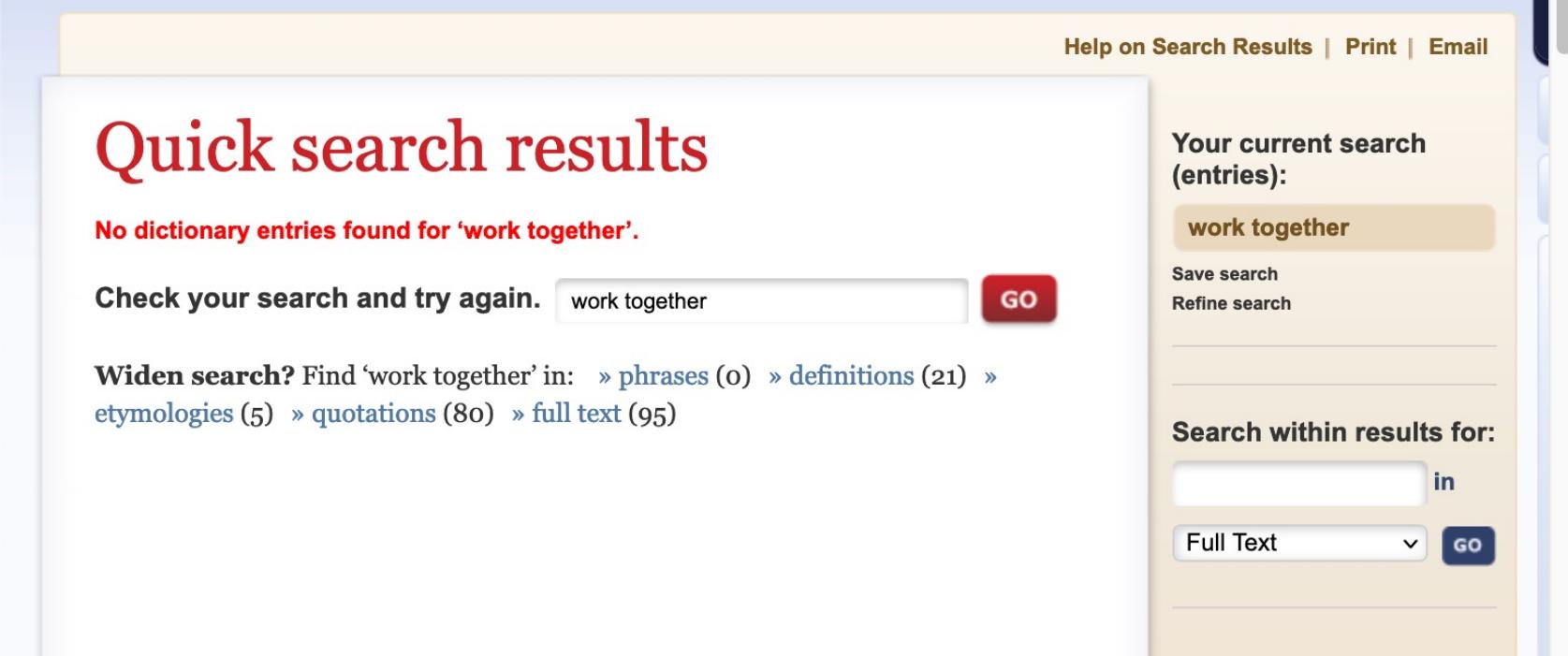
Check your search and try again.  GO

Widen search? Find 'work together' in: » phrases (0) » definitions (21) » etymologies (5) » quotations (80) » full text (95)

Your current search (entries):  
**work together**

Save search  
Refine search

Search within results for:  
 in  
Full Text GO



Advanced Search : Oxford English Dictionary +

https://www-oed-com.login.ezproxy.library.ualberta.ca

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

## Advanced search

Help on Advanced Search

Search in: Entries Senses Quotations

Search the text of OED entries:

Term

work in Definition

Additional Term

Near together in Definition Remove row

+ Add row

Options for NEAR/NOT NEAR

Five Words

Ordered

Reset Search

This screenshot shows the 'Advanced search' page of the Oxford English Dictionary. At the top, there's a 'Quick search' bar with a 'GO' button. Below it, links for 'Lost for Words?', 'Advanced search', and 'Help' are visible. The main title 'Oxford English Dictionary' is displayed with its subtitle 'The definitive record of the English language'. A large red header 'Advanced search' is centered on the page. Below it, there are tabs for 'Search in: Entries', 'Senses', and 'Quotations', with 'Entries' currently selected. A section titled 'Search the text of OED entries:' contains two search fields. The first field has 'work' in the 'Term' input and 'Definition' in the 'in' dropdown. The second field has 'near' in the 'Near' dropdown, 'together' in the 'Additional Term' input, and 'Definition' in the 'in' dropdown. There are 'Remove row' and '+ Add row' buttons for the second field. To the right of these fields are 'Options for NEAR/NOT NEAR' settings, including a dropdown for 'Five Words' and a checked 'Ordered' checkbox. At the bottom right of the search area are 'Reset' and 'Search' buttons.

Advanced search results : Oxford +

ABOUT COMMUNITY BLOG Access: University of Albe

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

## Advanced search results

Showing 1-20 of 24 results in 24 entries

View as: List | [Timeline](#) Jump to alphabetical point:  GO [Entry](#) | [Frequency](#) | [Date](#)

1. agree, v. [View full entry](#) a1413  
...s, behaviour, etc.; to live or work together without conflict. A... ••••••••

2. bayanihan, n. [View full entry](#) 1958  
...ich the members of a community work together to accomplish a dif ••••••••

Your current search (entries):

Search term X  
**work in Definition**

Search term X  
**together in Definition**

Options for NEAR/NOT NEAR:  
Distance: Five Words

[Save search](#)  
[Refine search](#)

Advanced search results : Oxford +

← → C ⌂ ⌂ https://www-oed-com.login.ezproxy.library.ualberta.ca ☆ ☰ A B+ ⌂ ⌂ ⌂

6. co-operate, v. [View full entry](#) 1604  
...erārī to work. intransitive. To work together, act in conjunction...

7. co-work, v. [View full entry](#) 1613  
...rks co- 1 + work intransitive. To work together; to co-operate.1613 S....

8. get, v. [View full entry](#) c1175  
...ansitive. To cause (a crew) to work together.1843 W. F. MacMichae...

9. happy, adj. and n. [View full entry](#) a1387  
... ships a ship on which the crew work together harmoniously; figur...

10. knead, v. [View full entry](#) c950  
...ver, and pressing or squeezing together ; esp. to

in  
Full Text GO

Advanced Search : Oxford English Dictionary +

https://www-oed-com.login.ezproxy.library.ualberta.ca

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

# Advanced search

Help on Advanced Search

Search in: **Entries** Senses Quotations

Search the text of OED entries:

**Term**  
work in Definition

**Additional Term**  
Near with in Definition Remove row

**Options for NEAR/NOT NEAR**  
Five Words  Ordered

**Additional Term**  
Near another in Full Text Remove row

This screenshot shows the 'Advanced search' page of the Oxford English Dictionary. The main search term is 'work', set to search in 'Definition'. An additional term 'with' is added near 'work'. Another term 'another' is added near 'with', with the search scope expanded to 'Full Text'. The search interface includes tabs for 'Entries', 'Senses', and 'Quotations', and a 'Reset' button. On the right, there are options for 'NEAR/NOT NEAR' searches, currently set to 'Five Words' and 'Ordered'. A 'Help on Advanced Search' link is available at the top right.

Advanced search results : Oxford +

ABOUT COMMUNITY BLOG Access: University of Alber

OED | Oxford English Dictionary *The definitive record of the English language*

Quick search:  GO

Lost for Words? | Advanced search | Help

Help on Search Results | Print | Email

## Advanced search results

Showing 1-10 of 10 results in 10 entries

View as: List | [Timeline](#) Sort by: Entry | [Frequency](#) | [Date](#)

1. [bevel, adj. and n.<sup>1</sup>](#) [View full entry](#) 1562  
... frustum of a cone, is oblique with the axis, ... used to work in co... 

2. [collaborate, v.](#) [View full entry](#) 1871  
...ur): see -ate. intransitive. To work in conjunction with another ... 

Your current search (entries):

Search term [x](#)  
work in Definition

Search term [x](#)  
with in Definition

Search term [x](#)  
another in Full Text

Options for NEAR/NOT NEAR: [x](#)  
Distance: Five Words

## Useful (on-line) bi-lingual dictionary:

Finding results for all types of words and phrases,

E.g.

*collaborate, procrastinate, cellphone,  
work together*

No Search Results : Oxford Engl X collaborate | translate English to X

← → C ⌂ https://dictionary.cambridge.org/dictionary/eng Dictionary Translate Log in / Sign up Search

collaborate x English–French :

collaborate

f t

# collaborate

verb

/kə'læbəreɪt/

+ Ⓜ

**to work together (with someone) on a piece of work**

**collaborer**

- *He and his brother collaborated on a book about aeroplanes/airplanes.*

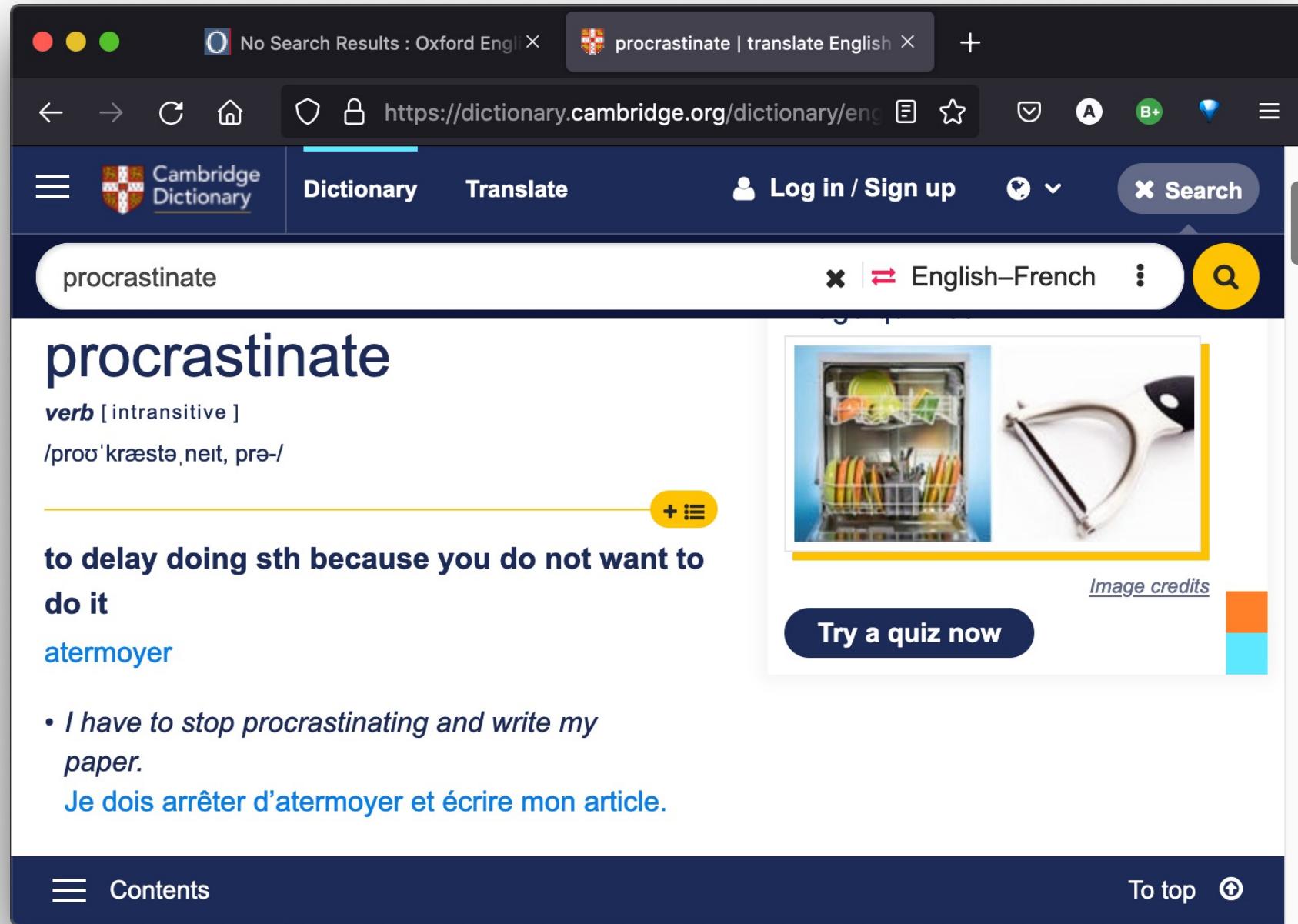
image quizzes

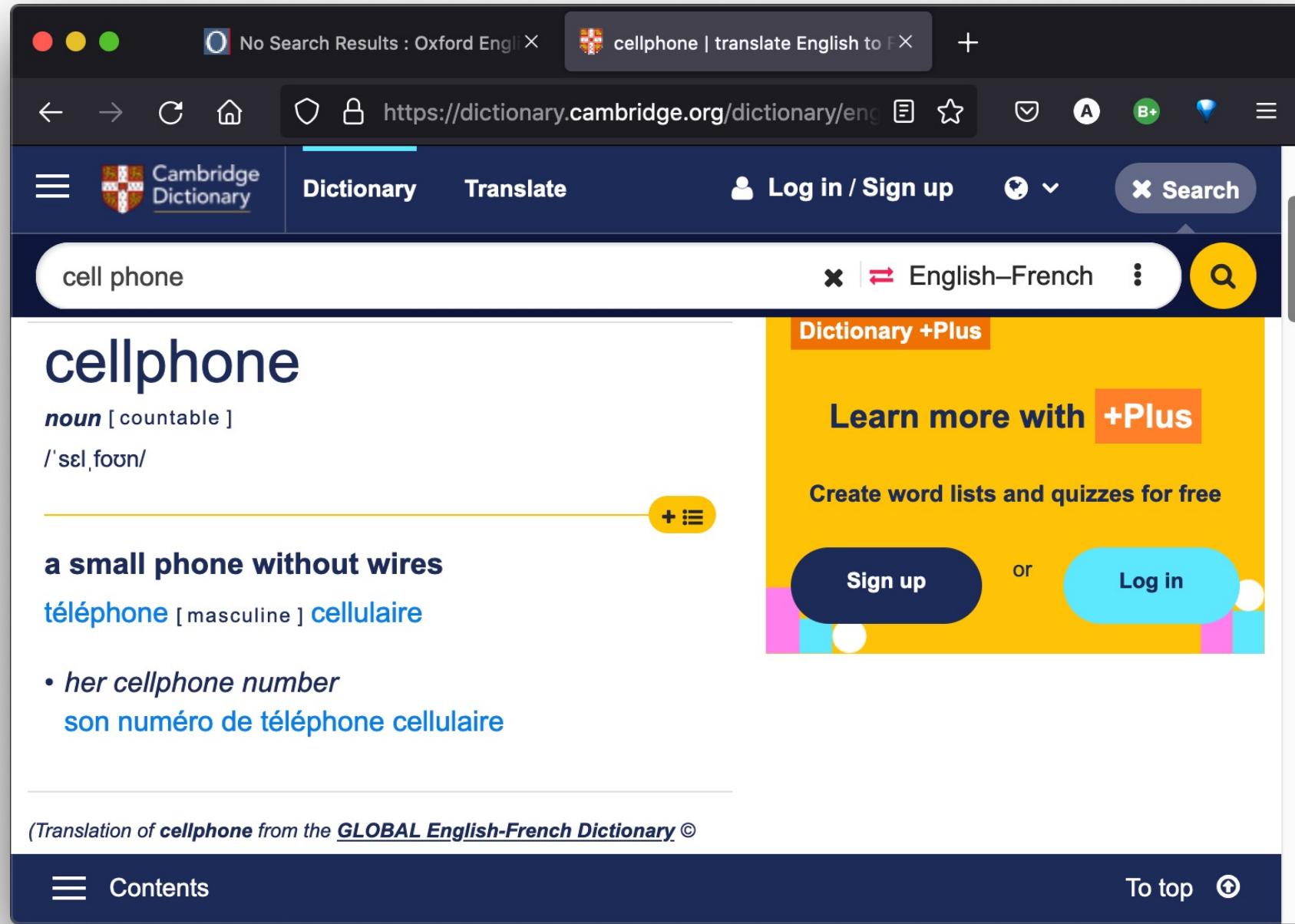


Image credits

Try a quiz now

☰ Contents To top ⓘ





No Search Results : Oxford Engl X work together - Did you spell it c X +

← → C ⌂ https://dictionary.cambridge.org/spellcheck/english ☆ ⌂ A B+ ⌂ ⌂ ⌂

☰ Cambridge Dictionary Dictionary Translate Log in / Sign up English–French ⌂ ⌂ ⌂

work together x English–French ⌂ ⌂ ⌂

## Search suggestions for **work together**

We have these words with similar spellings or pronunciations:

- [go together](#)
- [alone together](#)
- [stick together](#)
- [band together](#)
- [club together](#)

Test your vocabulary with our fun image quizzes



Image credits

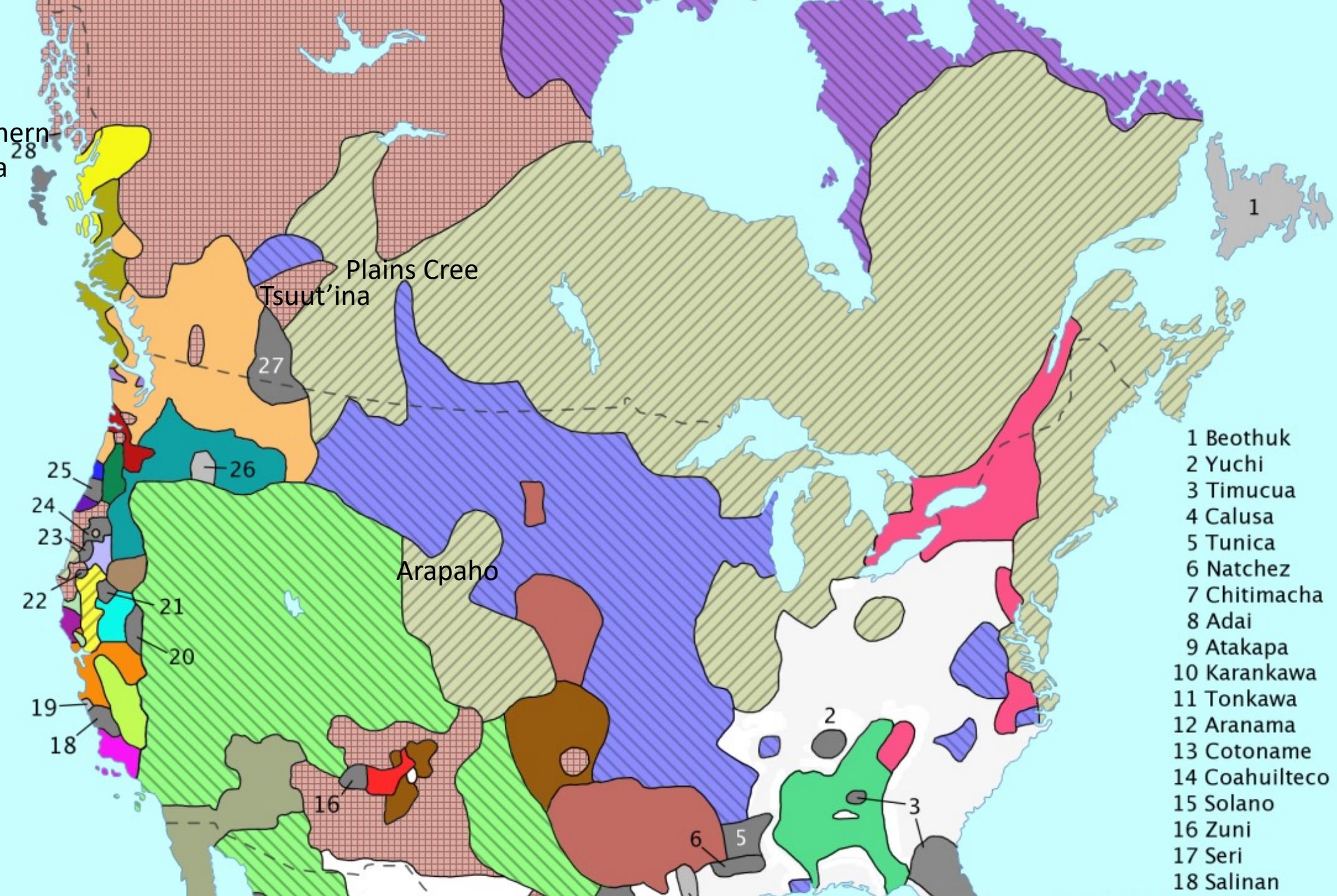
Try a quiz now

# Indigenous languages in focus ← North America

- Plains Cree (ISO: crk)
  - Algonquian
  - several thousands of speakers in Alberta, Saskatchewan, and the Canadian Western Plains
- Arapaho (ISO: arp)
  - Algonquian
  - several hundred speakers in Colorado and Wyoming (U.S.)
- Tsuut'ina (ISO: srs)
  - Dene
  - ~20 speakers in Southern Alberta
- Northern Haida (ISO: hdn)
  - Isolate
  - <10 speakers in the Haida Gwaii Islands off the coast of British Columbia

# Indigenous languages in focus ← North America

- Plains Cree (ISO: crk)
  - <https://itwewina.altlab.app>
- Arapaho (ISO: arp)
  - <https://nihii-tono.altlab.dev>
- Tsuut'ina (ISO: srs)
  - <https://gunaha.altlab.dev>
- Northern Haida (ISO: hdn)
  - <https://guusaaw.altlab.dev>



# Bilingual dictionaries for Indigenous languages (to English)

- Plains Cree / nêhiyawêwin ← Arok Wolvengrey (First Nations U)
  - *Cree : Words / nêhiyawêwin : itwêwina* (CW)
  - 23k+ entries
  - additional content from Maskwacîs Cree Dictionary (9k entries)
- Arapaho ← Andy Cowell (Colorado, Boulder)
  - *Arapaho Dictionary*
  - ~25k entries
- Northern Haida ← Jordan Lachler (UAlberta)
  - *Dictionary of Alaskan Haida*
  - ~5.5k entries (mostly verbs)
- Tsuut'ina ← Elder Bruce Starlight (Chris Cox) ← Johnny Onespot
  - *Onespot-Sapir Glossary* (OS)
  - ~12.5k entries (many inflected wordforms)

# Survey of 284 published dictionaries and lexical databases (Hieber, 2021)

- mean number of entries per dictionary/language: 5,772
- median number of entries: 4,321
- only 39 sources contain more than 10,000 entries
- only 5 have more than 20,000
- Only 2 sources reach 50 000 entries
  - Mundari (Glottocode: mund1320; ISO: unr)
  - Marwari (Glottocode: raja1256; ISO: mwr)

# Ball-park vocabulary sizes for majority language dictionaries

- All-encompassing dictionaries of any word in any register at any time
  - 500k+ entries
  - Oxford English Dictionary, Svenska Akademiens Ordbok
- Comprehensive dictionaries of essential vocabulary
  - 100-200k entries
  - Svenska Akademiens Ordlista, Kielitoimiston sanakirja
- Learner's dictionaries
  - 35k+ entries
  - Cambridge Learner's Dictionary of English

## Less-resourced/endangered languages

Finding results for rarer or more specialized terms  
is more challenging

E.g.

*collaborate, procrastinate, cellphone, blackberry,  
nokia*

No Search Results : Oxford Engl X Plains Cree: online dictionary: H X

← → C ⌂ 🔍 https://dictionary.plainscree.atlas-ling.ca/#/results ☆ ☈ A B+ ⚡ ⌂

kwe' waachiya boozhoo Tan-əkahk?  
hello kuei kwei boozoo  
tanisi! bonjour oki Taanshi?

# nēhiyawēwin

## Plains Cree dictionary

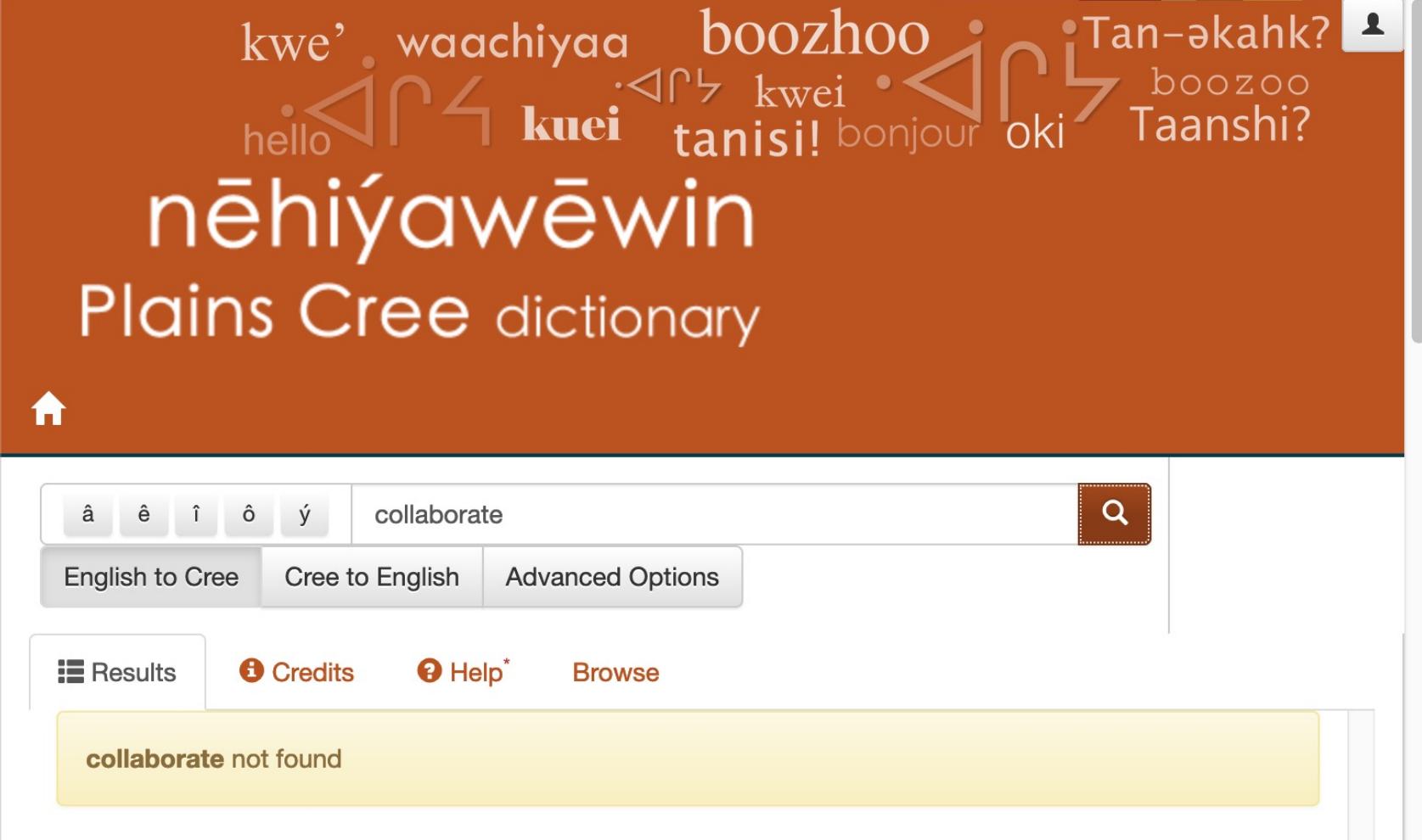
Home

â ê î ô ý collaborate

English to Cree Cree to English Advanced Options

Results Credits Help\* Browse

collaborate not found



No Search Results : Oxford Engl X Plains Cree: online dictionary: H X

← → C ⌂ 🔍 https://dictionary.plainscree.atlas-ling.ca/#/results ☆ ☈ A B+ ⚡ ⌂

kwe' waachiya boozhoo Tan-əkahk?  
hello kuei kwei boozoo  
nēhiyawēwin Taanshi?

# nēhiyawēwin

## Plains Cree dictionary

Home

â ê î ô ý procrastinate

English to Cree Cree to English Advanced Options

Results Credits Help\* Browse

procrastinate not found

The screenshot shows a web browser window with the address bar displaying the URL <https://dictionary.plainscree.atlas-ling.ca/#/results>. The main content area is titled "nēhiyawēwin Plains Cree dictionary". At the top, there's a decorative banner with several Cree words and their English meanings: "kwe'" (hello), "waachiya" (hello), "boozhoo" (hello), "Tan-əkahk?" (How are you?), "kuei" (yes), "kwei" (yes), "boozoo" (yes), and "Taanshi?" (What's your name?). Below the banner, the title "nēhiyawēwin" and "Plains Cree dictionary" is displayed. A search bar contains the word "procrastinate" and includes a phonetic key for Cree vowels (â, ê, î, ô, ý). Below the search bar are three tabs: "English to Cree", "Cree to English", and "Advanced Options". At the bottom, there are links for "Results", "Credits", "Help\*", and "Browse". A yellow message box at the bottom states "procrastinate not found".

No Search Results : Oxford Engl X Plains Cree: online dictionary: H X

← → C ⌂ 🔍 https://dictionary.plainscree.atlas-ling.ca/#/results ☆ ☈ A B+ ⚡ ⌂

kwe' waachiya boozhoo Tan-əkahk?  
hello kuei kwei boozoo  
nēhiyawēwin Taanshi?

# nēhiyawēwin

## Plains Cree dictionary

Home

â ê î ô ý work together

English to Cree Cree to English Advanced Options

Results Credits Help\* Browse

work together not found

The screenshot shows the homepage of the Plains Cree online dictionary. The top navigation bar includes tabs for 'Oxford Engl' and 'Plains Cree: online dictionary: H'. The main content area has a brown background with white text. It displays several Cree words with their English meanings: 'kwe'' (hello), 'waachiya' (hello), 'boozhoo' (hello), 'Tan-əkahk?' (hello), 'kuei' (hello), 'kwei' (hello), 'boozoo' (hello), and 'Taanshi?' (hello). Below this, the title 'nēhiyawēwin' is displayed in a large, stylized font, followed by 'Plains Cree dictionary'. A search bar at the top has the text 'work together' entered. Below the search bar are three buttons: 'English to Cree', 'Cree to English', and 'Advanced Options'. At the bottom of the page, there are links for 'Results', 'Credits', 'Help\*', and 'Browse'. A yellow banner at the very bottom says 'work together not found'.

No Search Results : Oxford Engl X Plains Cree: online dictionary: H X

https://dictionary.plainscree.atlas-ling.ca/#/results

blackberry

English to Cree Cree to English Advanced Options

Results Credits Help\* Browse

## Keywords 2

▶ blackberry ENG 4

- ◀Γ^nđΓ^ amiskomin NA-1  
black currant; yellow blackberry; literally: "beaver-berry"
- ◀Γ^nđΓòđ"Ω amiskominâhtik NI-3  
yellow blackberry tree; literally: "beaver-berry bush"
- ▶ b^nρUΓ^ kaskitêmin NI-1  
blackberry

No Search Results : Oxford Engl X Plains Cree: online dictionary: H X

← → C ⌂ 🔍 https://dictionary.plainscree.atlas-ling.ca/#/results ☆ ☈ A B+ ⚡ ⌂

kwe' waachiya boozhoo Tan-əkahk?  
hello kuei kwei boozoo  
nēhiyawēwin Taanshi?

# nēhiyawēwin

## Plains Cree dictionary

Home

â ê î ô ý nokia

English to Cree Cree to English Advanced Options

Results Credits Help\* Browse

nokia not found

This screenshot shows the homepage of the Plains Cree online dictionary. The main title is 'nēhiyawēwin Plains Cree dictionary'. Above the title, there's a decorative banner with several Cree words and their English translations: 'kwe'' (hello), 'waachiya' (hello), 'boozhoo' (hello), 'Tan-əkahk?' (How are you?), 'kuei' (yes), 'kwei' (yes), 'boozoo' (yes), and 'Taanshi?' (What's your name?). Below the title, there's a search bar with the word 'nokia' typed in. Underneath the search bar are three buttons: 'English to Cree', 'Cree to English', and 'Advanced Options'. At the bottom of the page, there's a yellow banner with the text 'nokia not found'.

# Even the most frequent English lemmata are not always found in the definitions

Top Lemmata	Plains Cree	Arapaho	Northern Haida	Tsúùt'ínà
100	99	100	93	91
200	194	198	174	145
300	287	295	249	197
400	374	393	315	237
500	462	486	378	280
600	554	578	441	325
700	639	668	494	348
800	719	759	543	385
900	809	853	598	418
1000	880	939	641	447

Any wordforms corresponding to the top lemmata in Corpus of Contemporary American English (COCA)

# Missing high frequency English words (all 4 dicts)

Rk(all)	Rk(n/v/j/r)	WORD	Rk(all)	Rk(n/v/j/r)	WORD
• 168	265	percent	• 807	971	detail
• 206	310	national	• 853	1020	Democratic
• 288	406	policy	• 891	1059	politics
• 299	417	data	• 912	1081	user
• 368	494	thanks	• 931	1102	attorney
• 481	616	international	• 934	1107	budget
• 538	680	author	• 967	1144	senator
• 595	744	administration	• 977	1155	Senate
• 632	784	sir	• 978	1156	violence
• 644	796	career	• 991	1171	civil
• 678	830	candidate	• 1009	1190	institution
• 696	850	meeting	• 1666	1872	review
• 724	882	network	• 2442	2674	professional
• 727	885	district	• 4280	4559	left
• 737	896	theory	• 4636	4925	campaign
• 792	956	reality			

“Missing” words: Government, legislature, and finance + modern technology + abstract concepts

More specific words such as **collaborate** or **procrastinate** are much further down in terms of frequency

# Reasons for smaller dictionary sizes

- many documentary projects are limited by time and money
- many indigenous languages also suffer from lexical attrition ← remaining speakers may simply not remember as many words as their predecessors once did
- number of lexemes may in fact be smaller than speakers of major Indo-European languages are used to
  - Words may cover a broader semantic field
  - Productive derivation based on a smaller number of roots

# Challenges for vocabulary/dictionary search

## Language-wise

- language may not have a specific term for the concept the user is searching for; and
- language has a term for the search query, but no definition exactly matches that query.

## Lexicography/dictionary-wise

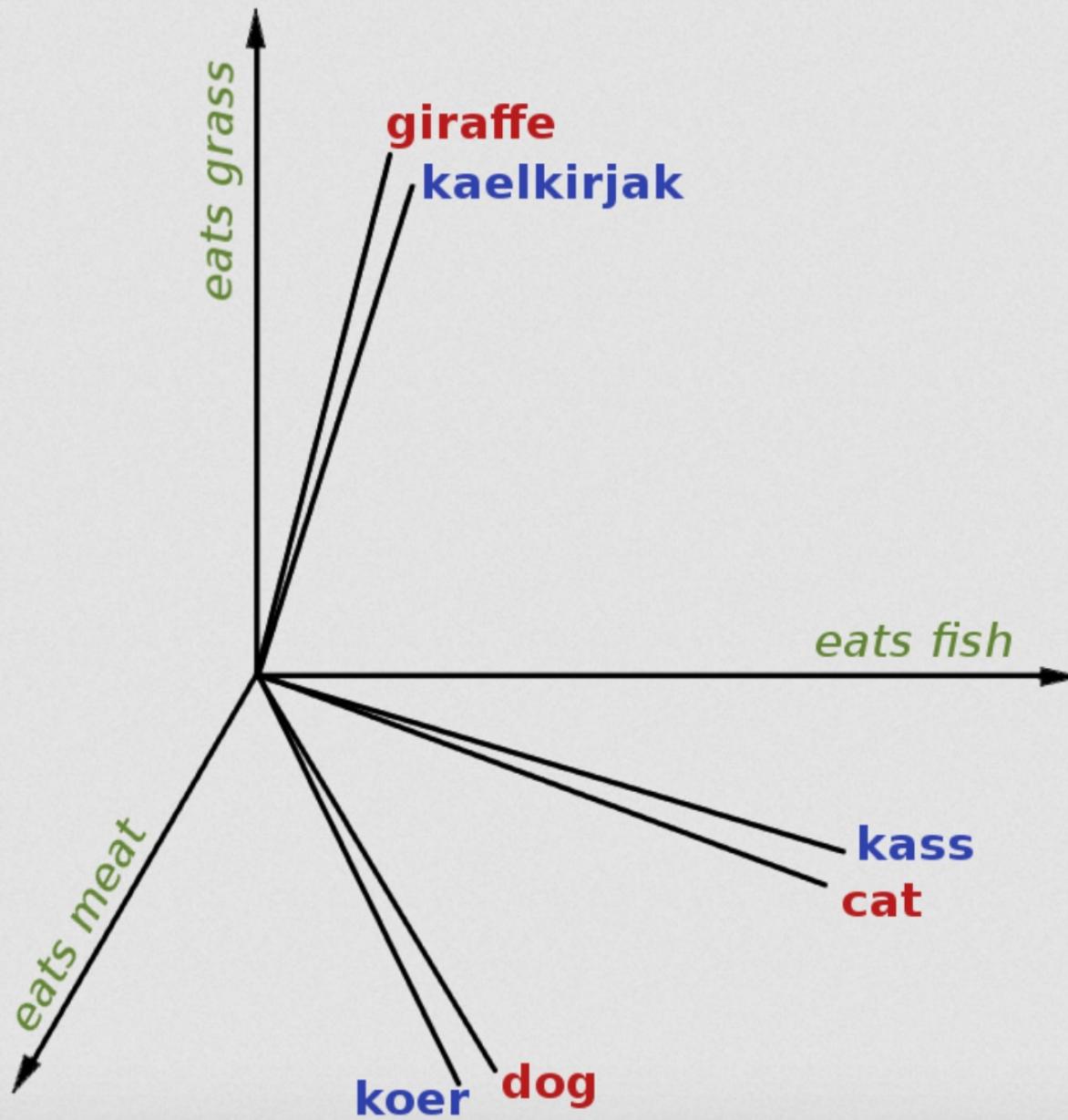
- definitions do not encompass the entirety of the semantic breadth of a term
- focus on basic level terms to the exclusion of others
- intentional rather than extensional definitions
- idiomatic meanings might not be explicitly included

# (Some) previous solutions

- expanding the search target space words with synonyms
  - and possibly with other lexical relations
- restricted to finding a match with the search word(s) and the target words (in definitions) via the synonyms
  - categorical in nature
  - either there is a match or there is not – there are no partial/gradient matches when comparing search and target words

# Our solution for English-to-X search – pre-trained embeddings

- make use of pre-trained 300-dimensional word vectors (“embeddings”) for English
  - word2vec (Mikolov et al. 2011) + Google News corpus ← 100? billion word tokens
  - 3 million word vectors
  - covers many orders of magnitude more central English vocabulary than the set of word tokens in the English definitions in any of the dictionaries
- create vectors for all English definitions by averaging the individual vectors of all the content words
- create a vector for English search terms similarly
- rank similarity of dictionary entries with search terms based on the cosine distance of the associated (aggregate) vectors



## mâmawohkamâtowak ⓘ

👤 ➡ like: nipâw 📖

1. they do things together, they cooperate cw
2. they work (at it/him) together as a group cw

## mihcêtohkamawêwak ⓘ

👤 ➡👤 like: nitonawêw 📖

1. they work together to help s.o. cw

## wîcihitowak ⓘ

👤 ➡ like: nipâw 📖

1. they help one another, they cooperate with one another cw

## mâmawêyatiwak ⓘ

👤 ➡ like: nipâw 📖

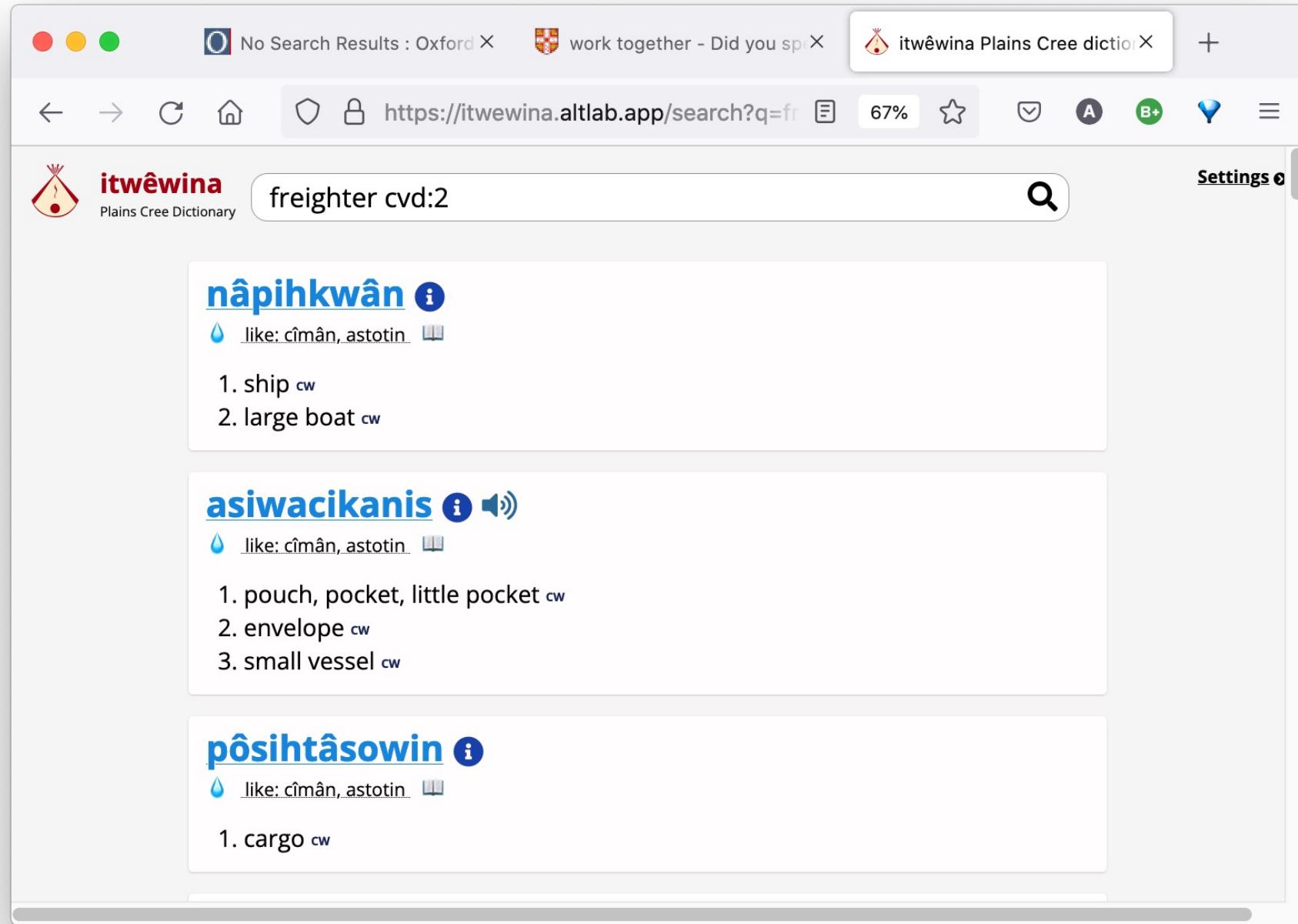
1. they go together, they proceed together cw
2. they are assembled, they are together as a group, they are together in numbers cw

# Advantages

- works even if there are no shared words between the search terms and the definitions
  - vectors are compared, not individual words
  - requires that at least one search term and at least one word in the definitions must be among the GNews vectors
- (almost) “everything” can be compared on a gradient scale
- works with any bilingual dictionary where a less-resourced endangered/Indigenous language is paired with a majority language
  - for which vector embeddings already exist, or
  - can be trained due to available extensive corpora
- semantically similar or related matches are presented alongside more exact matches

# Results - examples

- *osâwastotin*: yellow hat
  - **nîpâmâyâtastotin**: purple hat
  - **astotin**: hat, cap, headgear
  - **osâwêkin**: yellow material, yellow cloth
  - **osâwasâkay**: yellow dress, coat





No Search Results : Oxford



work together - Did you sp



collaborate cvd:2 — itwēwina



https://itwewina.altlab.app/search?q=collaborate



67%

itwēwina  
Plains Cree Dictionary

collaborate cvd:2



Settings

## mâmawohkamâtowak ⓘ

like: nipâw

1. they do things together, they cooperate cw
2. they work (at it/him) together as a group cw

## mihcêtohkamawêwak ⓘ

like: nitonawêw

1. they work together to help s.o. cw

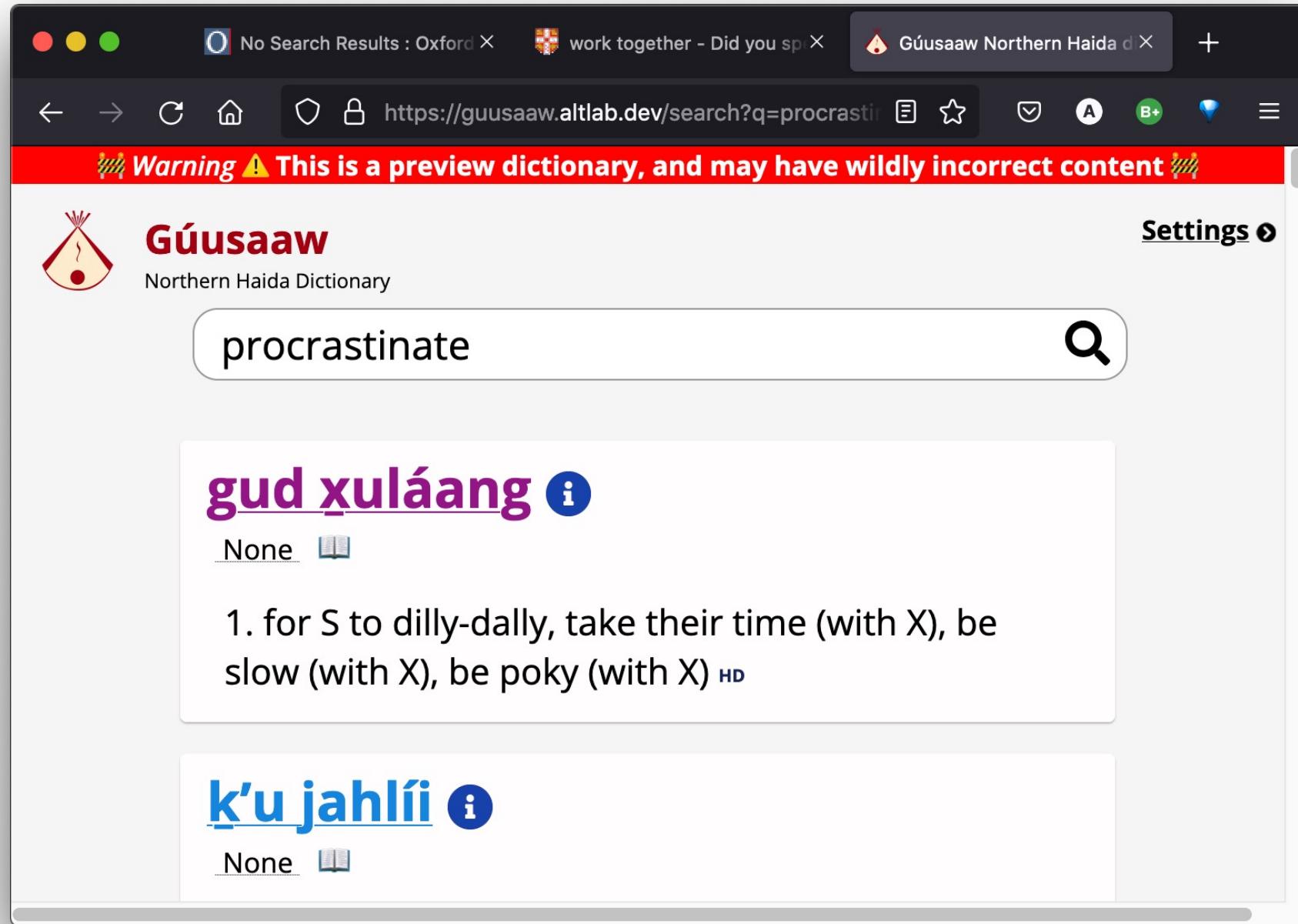
## wîcihitowak ⓘ

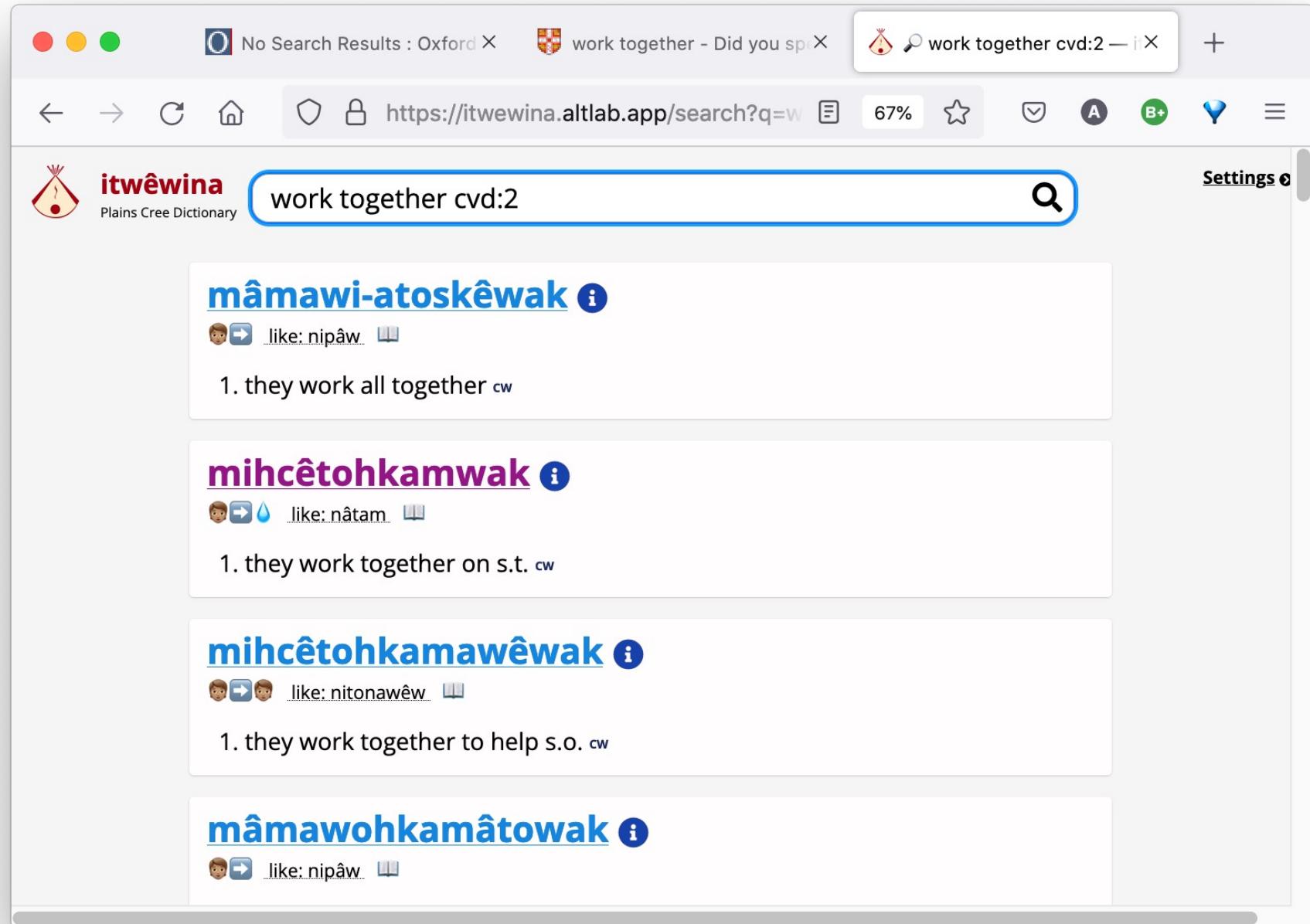
like: nipâw

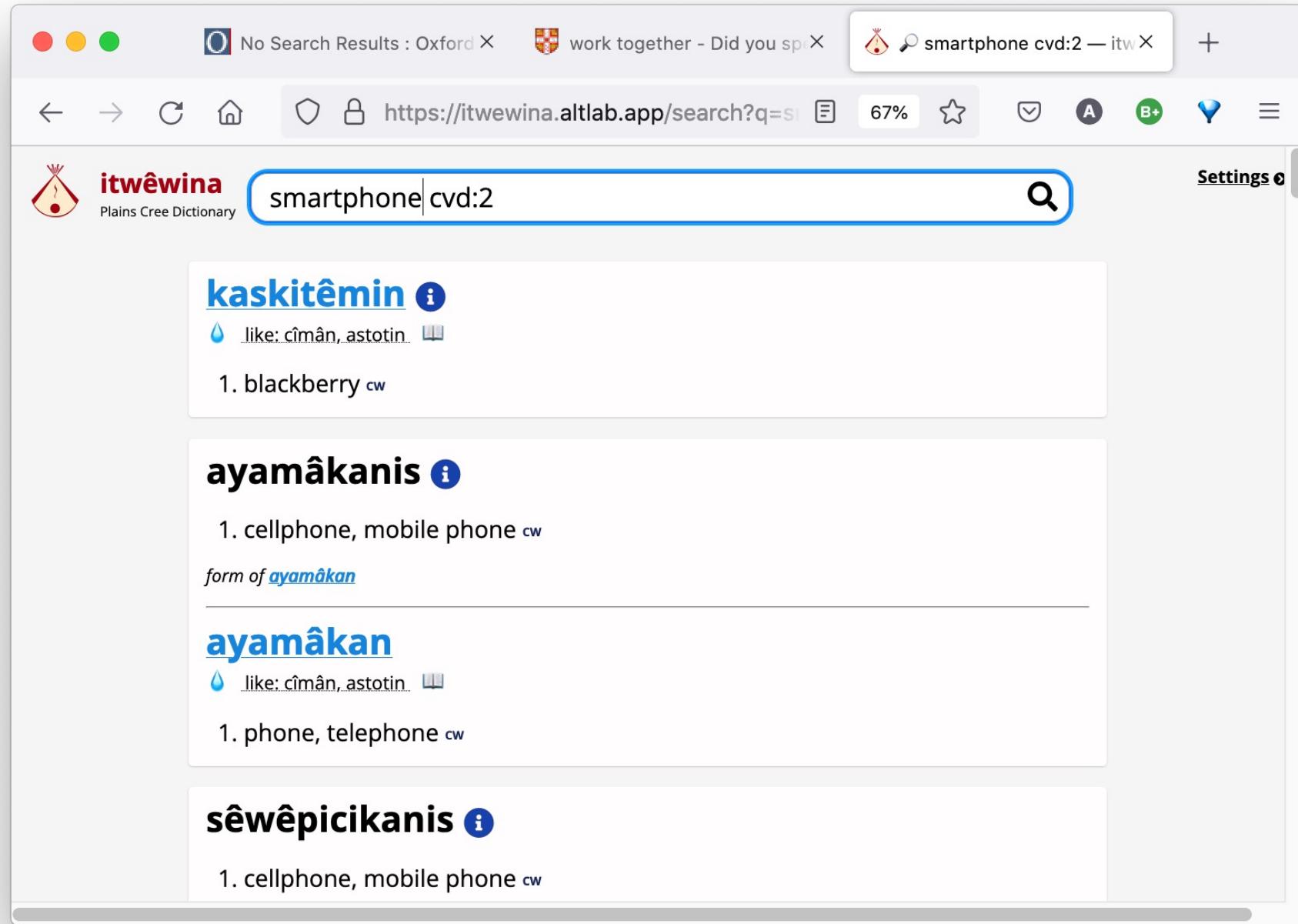
1. they help one another, they cooperate with one another cw

## mâmawêyatiwak ⓘ

like: nipâw







No Search Results : Oxford X work together - Did you sp X itwêwina Plains Cree dictio X +

← → C ⌂ ⌂ https://itwewina.altlab.app/search?q=n ↻ 67% ⭐ ⓘ B+ ⚡ ⌂ ⌂

 **itwêwina** Plains Cree Dictionary nokia cvd:2 

**sêwêpitamawêw** ⓘ

like: nitonawêw 

1. s/he phones s.o., s/he calls s.o. by phone cw
2. s/he makes (it) ring our for s.o. cw
3. He phones him. MD

**sêwêpitam** ⓘ

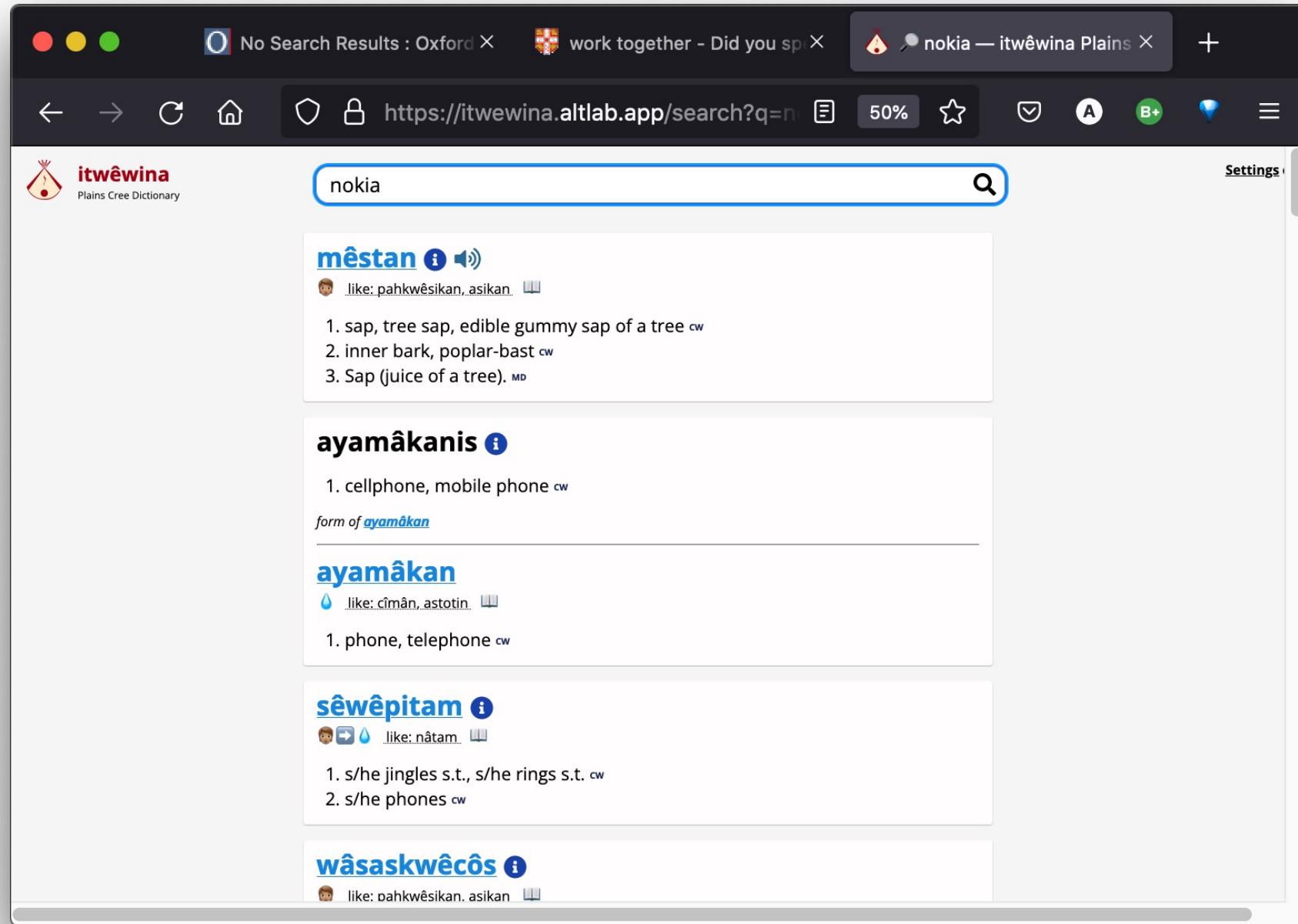
like: nâtam 

1. s/he jingles s.t., s/he rings s.t. cw
2. s/he phones cw

**wâskwêcowas** ⓘ

like: pahkwêsikan, asikan 

1. apple cw



# Results – qualitative → quantitative evaluation

- 26 high-frequency English lemmata not found in any of the dictionaries
  - top 10 cosine-vector-distance ranked dictionary entries
- Qualitative scales
  - high-quality result
  - moderate-quality result
  - poor result

168 265 percent										
	<i>nisto-sôniyâs</i>	<i>nîso-sôniyâs</i>	<i>pêyak-sôniyâs</i>	<i>mitâtahto-piwanîk</i>	<i>mâmâsîs</i>	<i>sôniyâs</i>	<i>mihkokwaniy</i>	<i>âpihcawês</i>	<i>kêkâc</i>	<i>kosikwan</i>
206 310 national	<b>pîkaniwiyinînâhk</b>	<i>nâtakâm</i>	<i>pêyakôskân</i>	<b>okimânâhk</b>	<i>nâtakâsi-</i>	<b>nêhiyawîhtwâwin</b>	<b>nahkawiyinînâhk</b>	<b>nêhiyawîhtwâw</b>	<b>kihci-okimânâhk</b>	<i>isiwêpan</i>
288 406 policy	<b>wiyasiwêwin</b>	<b>kakêskwêwin</b>	<i>oyasiwêwin</i>	<i>kwayaskowêwin</i>	<b>oyikâpawiw</b>	<b>kwêskasiwêwin</b>	<b>kwêskasiwêw</b>	<i>manâcimâkan</i>	<i>ostêsimâwasinahikan</i>	<i>ostêsimâwasinahikan</i>
299 417 data	<b>kiskêyihtâsowin</b>	<i>kwayaskomîwêwin</i>	<b>âcimostâkêwin</b>	<b>kiskimanikhêwin</b>	<i>itâcimowin</i>	<i>tipahaskêw</i>	<b>oyakihcikêwin</b>	<i>tâsaham</i>	<i>kiskiwêhikêmakan</i>	<i>mâwasakopêham</i>
481 616 international	<i>opîtatowêw</i>	<i>nêhiyawî-</i>	<i>nêhiyawî-</i>	<b>pîtosi-</b>	<i>pisikiw</i>	<i>pîtatowêw</i>	<i>iyiniw</i>	<i>môniyâw</i>	<i>nêhiyawî-wîhowin</i>	<i>môniyâs</i>
538 680 author										
	<b>masinahikan</b>	<b>nîkânitwêwin</b>	<i>kihci-masinahikan</i>	<i>nêhiyawasinhikan</i>	<i>nêhiyawî-masinahikan</i>	<b>otâcimow</b>	<b>masinahikêw</b>	<i>atâwêsinahikan</i>	<i>naspasinhikêw</i>	<b>masinahikêwin</b>
595 744 administration	<b>kihci-okimâw</b>	<b>okimâwiwin</b>	<i>maskominânâhtik</i>	<i>okimâwikamik</i>	<i>cahkâskwêyâw</i>	<b>okimânâhk</b>	<i>nâtakâsi-</i>	<i>nâtakâm</i>	<i>kaskitêminânâhtik</i>	<i>okimâhkâniwiw</i>
644 796 career	<i>ispîhtaskîwin</i>	<i>kîsatoskêw</i>	<i>pimâtisítotam</i>	<i>itâcîhow</i>	<i>nâtamâkêw</i>	<i>pimâtisiwin</i>	<i>kakâyawâcîhow</i>	<i>kiskêyihtam</i>	<i>nikamoham</i>	<i>kwêskâtisiw</i>
678 830 candidate	<b>omawinêhikêw</b>	<b>pimipahtâwin</b>	<b>nawasônîkêwin</b>	<b>omawinêskâkêw</b>	<b>onawasônîkêw</b>	<i>nôtinâkan</i>	<i>okimâw</i>	<i>nawasônîkêw</i>	<i>nawasônîwêw</i>	<b>otâpapîstamâkêw</b>
724 882 network										
	<i>pamîstâkêwin</i>	<i>kosâpahcikan</i>	<i>tahkoskêwin</i>	<i>têhtahokan</i>	<i>âpacihcikan</i>	<i>waskawîwin</i>	<b>âniskômohcikan</b>	<i>pamihwêwin</i>	<i>môniyâw-âpacihcikan</i>	<b>ayamâkanis</b>
727 885 district	<b>kihci-ôtêna</b>	<b>ôtêna</b>	<i>kiskinwahamâtowin</i>	<b>ôtêna</b>	<b>ôcénâs</b>	<i>kiskinwahamâkêwîkam</i>	<i>kiskinwahamâtôtâpân</i>			
						<i>ik</i>	<i>âsk</i>	<i>âwatawâsiswâkan</i>	<i>askiy</i>	<i>kiskinwahamâtowikam</i>
737 896 theory	<b>itêyihtâkan</b>	<b>tâpwéhtamowin</b>	<i>kwayaskwêyihtamowin</i>		<i>manitowêyihtamowin</i>	<i>omâyipayihcikêw</i>	<i>nâspici-wiyasiwêwin</i>	<i>mistakêyimisowin</i>	<i>cîsihowin</i>	<i>nahêyihtawêyihtam</i>
										<i>nihtâwêyihtamowin</i>
792 956 reality	<b>tâh-tâpwê</b>	<b>tâpwêwin</b>	<b>tâpwê</b>	<i>tâpwê_ësa</i>	<i>kosâpahcikan</i>	<i>pawâtaski</i>	<i>pawâtamowin</i>	<i>tâpwêmêw</i>	<i>maci-pawâmiw</i>	<i>tâpwêmakan</i>
853 1020 Democratic	<b>nawasônîkêwin</b>	<b>pimipahtâwin</b>	<i>owiysiwiwîkîmâw</i>	<b>onawasônîkêw</b>	<i>môcikihtâwin</i>	<i>mistikomin</i>	<i>mistiko-pakâ</i>	<i>omawinêhikêw</i>	<i>tahkiskikêw</i>	<i>maskominânâhtik</i>
891 1059 politics	<b>tahkiskikêw</b>	<b>mâmiskötamâkêwin</b>	<i>ayamihâwin</i>	<i>isihcikêwin</i>	<i>pimâchîhowin</i>	<i>mâmawinitowin</i>	<i>mitêwiwin</i>	<i>âhtisihcikêw</i>	<i>pimipahtâwin</i>	<i>nawasônîkêwin</i>
912 1081 user	<i>mamâhtâwiyiniw</i>	<i>owâpahcikêw</i>	<i>onitawâpênikêw</i>	<i>sêhkê-atoskêmakan</i>	<i>otayamihcikêw</i>	<i>sêhkêpayi</i>	<i>masinatahikêw</i>	<i>napopayi</i>	<i>apisîsiw</i>	<i>kâsîpayi</i>
931 1102 attorney	<b>owiysiwiw</b>	<b>oyasiwêiyiniw</b>	<b>opîkiskwêstamâkêw</b>	<b>kakêskikhêmowin</b>	<b>owiysiwiw</b>	<i>opaminikêw</i>	<i>masinahikêsis</i>	<i>onîkânapi</i>	<b>kakêskwêwin</b>	<i>okimâsis</i>
934 1107 budget	<i>otinamâsowin</i>	<b>mêstînikêw</b>	<i>manisam</i>	<i>kwayaskowêwin</i>	<i>nâtasinahikanêwin</i>	<i>kiskinwahamâtowin</i>	<i>kîskisam</i>	<i>masinahamâkêw</i>	<i>kaskihkotêw</i>	<i>kaskihkocikêw</i>
967 1144 senator	<b>owiysiwiwîkîmâw</b>	<i>okimâwikamik</i>	<i>nimanâcimâkan</i>	<i>ayamihêwiyinîwiw</i>	<i>iskwêw-atoskêyâkan</i>	<b>kihcapiwin</b>	<b>apiwin</b>	<i>omasinahikêsiswiw</i>	<i>nôtiskwêwêw</i>	<i>ayamihêwiyinîwiw</i>
977 1155 Senate	<b>owiysiwiwîkîmâw</b>	<b>mâmawapiwin</b>	<b>kwêskasiwêwin</b>	<i>maskominânâhtik</i>	<i>okimâwikamik</i>					
978 1156 violence										
	<b>nipahtâkêw</b>	<b>misi-maci-tôtamowin</b>	<b>otihtinikêw</b>	<i>otihtiniwê</i>	<b>ayiwâkêyimowin</b>	<b>misiwanâcihcikêw</b>	<i>macâtisiwin</i>	<b>otihtinikêw</b>	<b>nipahtâkêski</b>	<i>macitêhêwin</i>
991 1171 civil	<i>nôtinitowin</i>	<i>nôtinitikêw</i>	<i>mihkowi</i>	<i>paskêwihitowak</i>	<i>manâcimâkan</i>	<i>askiy</i>	<i>kîspin</i>	<i>mâyêyihcikêw</i>	<i>kihci-wîkihtowin</i>	<i>oyasiwêwin</i>
1009 1190 institution										
	<i>ayamihêwihkân</i>	<i>kiskinwahamâtowin</i>	<i>miscikos</i>	<i>okiskinwahamâkan</i>	<i>kiskinwahamâkosiwin</i>	<i>môniyâwi-ayamihâwin</i>	<i>otipêyimisow</i>	<i>mâmawinitowin</i>	<i>kiskinwahamâkêw</i>	<i>kihcîyiniw</i>
1666 1872 review	<b>oyakihcikêw</b>	<b>itasiwêwin</b>	<i>kakwêcihkêmowin</i>	<b>oyakihcikêw</b>	<b>kanawâpamisowin</b>	<i>naspasinhikêw</i>	<i>masinahikan</i>	<i>nisitohtamôhiwêwin</i>	<i>wiyîhcikêw</i>	<b>oyakihtamawêw</b>
2442 2674 professional	<b>nahtâ-</b>	<i>pamihwêwin</i>	<b>nihtâ-</b>	<i>nôhkam</i>	<b>kihcîyihkâsowin</b>	<i>tipiyawê</i>	<i>nimosôm</i>	<i>kêhtê-</i>	<i>nitôskinikîmis</i>	<i>nakacihtamohiwêwin</i>
4636 4925 campaign	<b>nawasônîkêw</b>	<b>niminahtâw</b>	<i>âhkamâvîmowin</i>	<i>nîkîhkâkâtam</i>	<i>nakasitowak</i>	<i>nîmohtêskanaw</i>	<i>tahkiskikêw</i>	<i>konica</i>	<i>konita</i>	<i>nîkonita</i>

168 265 percent	nónou3ó.o	cébeeníhi'	cébe'eini.ii'oo-	bébenéh	koh'u-	béeséis	kóuh'usí-	3eecei'iinoo-	3eecei'iinéíhi-	3eecé'iinííhi'
206 310 national	ciiskóihu-	wóóhonéníteeno'	wóóhoncé'eséítihího'	wóóhonítee-	céeciisikohúútooni'	bo'óóbe'	híteen	níhi'ikohúuhuu	nó'otéíhi-	tenéeteexáo3eit
288 406 policy	hóóxuwúut	ooxuwuut	hiinoxuwu-	neyóóxo'ée-	héítówo.o	cíino'onítóot	nooxúwu-	bóó'einéécee	bé'iicí3einí.iní_hé'iitowót	ní3ooxúwutí-
299 417 data	níitówo.o	héítowóot	he'íyo.o	3óóko3onkúutii-	bí'eenéet-	héítóbee-	heenéi'towóotííni-	cíh'o'wúuse3ó.o	3é.i.íyo.o	héítówo.o
481 616 international	wóóhoncé'eséítihího'	wóóhonéníteeno'	ni'óeenen	kóóxo'uusíínoo'oo'	kóóxo'óókuu'oo'	hee3éi'o'obee'	hotéisóó	héneesi-	heenéi'o'owúse'	nenebnih'oo3oo
538 680 author	hoo3íte.íhii	beise'íyei.it	beehíinísiit	wo3ónohéhiinóó-	wo3onohu-	woxusenini-	wo3ónohé.ihí-	ko3éíno3onohó.e	wo3onohó.e	wóóno3onohó.e
595 744 administration	niinoxuwut	nooh'úwunibíis	ko'éíno'ó.owu'	honíi'onibíis	yonáó3onó.uuyoo'	hóówoo'	biinonoo'óé	bóó'einéécee	bíis	cóó'oo'ee'
644 796 career	heenéecxooyéi'oo'	béetóu3e.i-	cécinib	híwo3onohóéni-	húúne'etíit	híne'etíit	kó.u3óó-	recordyih-	cowoxú.uní-	béénii'owúúni'
678 830 candidate	bíh'ihóox	beseyóótiini-	nii3kóóhu-	beséyeini-	see'etéét	niinoxuwut	cíito3onohoenee-	héénetíihii	neecéései	níitobéíhii
724 882 network	béi'icí3eiséénook	céi3wó.oo	céi3wó.onóooxunéeno'ó.owu'	wótotó.oo	woohnóócei.íse'	nonooxeihit	nooxéíhi'	nonoo'oeníh'ohúúnoo'	wó3onikúu3o.o	béi'hinooko3onohó.e
727 885 district	howóh'ó.owu'	neyéi3e.icebkóóhuu3éi hii	neyéi3e.íno'ó.owu'	neyéi3e.iníitobéíhii	niiniítowóoto'_neyei3éi no'ó.owu'	hécesneyéi3éíhii	co'ó.uutóu3eibéíhii	wóteenííhi'	neyéi3e.ínotii	neyéi3e.ísei
737 896 theory	yohowóh	hécesneyéi3einíitobéíhi i	heetéetoo	hóóxuwúut	héítówo.o	nihíit	hééteenííhi'	hí3einoonései	cenéetéénii'éhii	cíiyó.ootí-
792 956 reality	wó3onikúu3o.o	néenéé'owóbee-	hiixoohóó3í3e.i-	hí3oobéé-	héé3neenéé-	woxukónee-	noní3e'éexu-	bébeeyóot	kónee-	héé3neenóó-
853 1020 Democratic	see'etéét	bíh'ihóox	beseyóótiini-	notíkonít	nii3kóóhu-	hixoceib	nooh'úwunibíis	béesineníteeno'	beséyeini-	ko'éíno'ó.owu'
891 1059 politics	niicésiskúu3ei.it	cíino'onítóot	heesííne'etíi-	neenéísííne'etíi-	notíkonít	beseyóótiini-	béesineníteeno'	bíh'ihóox	híine'etíit	hóóko'óé-
912 1081 user	hotóonééhii	hotóoneehííni-	notíi3íye.íhii	heeneetoh-	bex	ce'íbes	neyéi3éíhii	hookú.u	iítee	inenitee
931 1102 attorney	nénebe'eebííhii	hooxúwuunén	hóóxuwúúsei	hóóxuwúúhuu	tó.ukú3e.íhii	wo'ótóu3e.íhii	heenéiseinéew-	cebko.ohúú3e.íhii	cee'íye.íhii	héé3neetóukú3e.íhii
934 1107 budget	hooxúún-	yóó3oyó.usíne-	neníixó.u'úsi'	cíiyó.o	hooxúúye.i-	yéíno.úh'u-	yóó3oyo.úh'u-	iixooxuun-	néeso.úh'u-	kooxésoot_séénookuu
967 1144 senator	bíh'ihóox	see'etéét	niinoxuwut	beníiinései	hóówoo'	biinonoo'óé	béteenih'oo3ó.uni-	nééceenibóoot	béteenih'óó3oo	cíihóúyeinééti-
977 1155 Senate	see'etéét	bíh'ihóox	cíihóúyeinééti-	ko'éíno'ó.owu'	yóó3oyó.usíne-	neníixó.u'úsi'	bé'iis	beseyóótiini-	cé'esbenenííhihi'	béé3eenetíítóóni-
978 1156 violence	wóowóxuh	nonóxoo3íhetíit	nóxowútoo-	3óo3owó.uyóot	hooxonooni-	nóxoni-	nóó'ooh-	hí3kúu3e.íhii	nóó'oohéíhii	nóónisoo'
991 1171 civil	hooxúwuunén	béexóó3etiít	tokóokeih'ínee-	notíkonít	be'ei'eenini-	be'ííni-	bóó'einéécee	ní'onéébetíit	bóó3etiiniíhi'	bóó3etiít
1009 1190 institution	koéínciyyó.ono_níítón onihéíhiinóú'u	tesco'ó.uutóu3éíhii	wo3onohóec'exó.owu'	wóóhono3ónohú-	bénení.iit	bení.iinén	niinoxuwut	tesco'ó.uutóu3eihííni-	cebico'ó.uto.u3éíno'ó.owu'	tesio'ó.uutóu3eibéíhii
1666 1872 review	kokouh'uh-	tééceenów-	kókoh'óéni-	nehyóhoníhi'	cooxúwutii-	kokóuh'ouh-	kokóuh'outíi-	kókoh'ó.uhú-	kokóuh'ó.úhetíit	kókoh'óuti-
2442 2674 professional	híwo3onohóéni-	nónono3itoo-	kóokúnoo'oo-(koo'kookunoo'oo?)	koto'u-	he'ííhi-	héeteenébé.ihíinóó-	heenéeteenebéíhiinóó-nihéyoo	he'íye.íhii	notónihe.ísei	
4636 4925 campaign	nii3kóóhu-	nééhiw-	beseyóótiini-	hiicoo'owon-	notíkonít	bóó3etiít	nonóúhetí-	nonóúhetihíhi-	nonóúh-	nonóúhu-

168 265 percent	<i>kwáadägaa</i>	<i>bíid_inuyáa</i>	<i>ka_dluwi</i>	<i>t'ánskad_hlgahl</i>	<i>ináasdlayaa</i>	<i>gu_kúnjuu</i>	<i>gud_gaad'agaa</i>	<i>hat'án_ináa</i>	<i>súlunang</i>	<i>sdahhunáng</i>
206 310 national	<i>nánt'as</i>	<i>gyáa_sahl_náa</i>	<i>k'án_gudáng</i>	<i>ga_gasgadáa</i>	<i>chalagúng</i>	<i>dángyaang</i>	<i>tláasala</i>	<i>ináa_t'asdliyáa</i>	<i>k'wáalsdliyaa</i>	<i>hláalgaa</i>
288 406 policy	<i>kéeng_hlangaa</i>	<i>da_k'asdliyáa</i>	<i>dáng_xánhlahldayaa</i>	<i>gi'uláa</i>	<i>sáagaa_tiiyáa</i>	<i>xaad_k'aláa</i>	<i>xánjuu</i>	<i>ki_tl'agáng</i>	<i>dáng'ulaa</i>	<i>gíihlagahl</i>
299 417 data	<i>k'wíidayaa</i>	<i>stla_k'wíidayaa</i>	<i>kíng_gudáng</i>	<i>k'ujgw_kaanáa</i>	<i>xaa_kaanáa</i>	<i>gangáa</i>	<i>keh_k'wíidayaa</i>	<i>hldiingáa</i>	<i>xi_k'wíidayaa</i>	<i>sgwáagaanaa</i>
481 616 international	<i>Xaat'áa</i>	<i>Jáabäniigaa</i>	<i>Git'áwyadaa</i>	<i>húnggangaang</i>	<i>k'áadaa</i>	<i>skáanst'ayaa</i>	<i>hláalgaa</i>	<i>lamduwáa</i>	<i>kaanáa</i>	<i>'wáahlayaa</i>
538 680 author	<i>ta_k'áalaang'_la'áaygaa</i>	<i>ta_k'áalang</i>	<i>st'ii_kéeng'_la'áaygaa</i>	<i>sgalangáa</i>	<i>táwk'aa'_la'áaygaa</i>	<i>stl'uwíi</i>	<i>kihl_gáajuwa</i>	<i>dagwáang</i>	<i>lableedgáa</i>	<i>k'aa'_la'áaygaa</i>
595 744 administration	<i>k'áadaa</i>	<i>'láanaa'_la'áaygaa</i>	<i>tajúu</i>	<i>hlk'awáng</i>	<i>xu_hlkílgangaang</i>	<i>sgaast'ayáa</i>	<i>gi'uláa</i>	<i>gi'adiigaa</i>	<i>gyáa_sahl_náa</i>	<i>gi'adaa</i>
644 796 career	<i>xidid</i>	<i>ja_dúu</i>	<i>xijuuláang</i>	<i>sgánhlahldayaa</i>	<i>tlakwiigáa</i>	<i>nats'ä_xíldayaa</i>	<i>náang_gulgáa</i>	<i>stl'a_skáawnang</i>	<i>kúnhlíyaa</i>	<i>gíihdayaa</i>
678 830 candidate	<i>dáalaa'_la'áaygaa</i>	<i>kihl_gáajuwa</i>	<i>'láanaa'_la'áaygaa</i>	<i>kínyaaw</i>	<i>gijgad_k'áwaang</i>	<i>ta_gináang</i>	<i>dlasdlí_láa</i>	<i>sku_k'áwaang</i>	<i>gyáa_sahl_náa</i>	<i>k'i_k'áwaang</i>
724 882 network	<i>gíihl</i>	<i>tlakhíldang</i>	<i>k'íiwii</i>	<i>sgáanuwaa</i>	<i>xawíi</i>	<i>dabwíi</i>	<i>k'ad_kingáang</i>	<i>tláasala</i>	<i>na_sk'iila</i>	<i>sgih'l'adáa</i>
727 885 district	<i>'láanaa'_la'áaygaa</i>	<i>gúngahlgwaay</i>	<i>tláasala</i>	<i>xáwtl'ädangaa</i>	<i>'láanaa_awyáa</i>	<i>siináa</i>	<i>k'ad_kingáangwaa</i>	<i>nánt'as</i>	<i>tl'ahlaláa</i>	<i>gin_sk'at'adáa'_la'áaygaa</i>
737 896 theory	<i>k'íigangaa</i>	<i>xaad_dlánsunaa</i>	<i>tlakmíits'angaa</i>	<i>áalaa</i>	<i>hlkáak'aa</i>	<i>kílmíits'angaa</i>	<i>gu_míits'angaa</i>	<i>kéeng_hlangaa</i>	<i>kyúu_tláayangaa</i>	<i>kihljúu</i>
792 956 reality	<i>kílsiid</i>	<i>xa_xángisälang</i>	<i>k'íigangaa</i>	<i>k'ah_xángisälang</i>	<i>k'áangahlt'as</i>	<i>gyaahländä_sgáanuwa</i>	<i>káast'ayaa</i>	<i>xaad_dlánsunaa</i>	<i>yahk'iyáa</i>	<i>k'ut'ahl</i>
853 1020 Democratic	<i>gáwt'unaa</i>	<i>'láanaa'_la'áaygaa</i>	<i>kaygud'áang</i>	<i>jíidaal</i>	<i>kihl_gáajuwa</i>	<i>dáng_xánhlahldayaa</i>	<i>giwaa</i>	<i>k'ihlă_tl'awáansdlayaa</i>	<i>k'ihlă_tl'aawáa</i>	<i>kéeng_hlangaa</i>
891 1059 politics	<i>xaad_k'aláa</i>	<i>kíl_xángisälang</i>	<i>kéeng_hlangaa</i>	<i>dlat'ayaa</i>	<i>sáagaa_tiiyáa</i>	<i>ináa</i>	<i>kihldáang_k'íisgadaa</i>	<i>k'áadaa</i>	<i>hlkwiidayáa</i>	<i>hlgáwyaaang</i>
912 1081 user	<i>dah</i>	<i>níijang</i>	<i>xáwla'ihl</i>	<i>gyáa_sgáw</i>	<i>iihliyaa</i>	<i>tl'ast'ä_káahliyaa</i>	<i>yáadgaa</i>	<i>ít'ajaa</i>	<i>xaad_k'aláa</i>	<i>gyúuwad</i>
931 1102 attorney	<i>láayaagaa</i>	<i>dáalaa'_la'áaygaa</i>	<i>'láanaa'_la'áaygaa</i>	<i>st'ii_kéeng'_la'áaygaa</i>	<i>ta_gíng_kílisäläang'_la'áaygaa</i>	<i>lableedgáa</i>	<i>ja_dáang</i>	<i>k'íijuwaa</i>	<i>hlgwäyáa</i>	<i>ja_dúu</i>
934 1107 budget	<i>sálda_k'íisvlaahl</i>	<i>'láanaa'_la'áaygaa</i>	<i>gyáa_sgáw</i>	<i>bíidgaa</i>	<i>gíi'adiigaa</i>	<i>gíi'adáa</i>	<i>bíid_inuyáa</i>	<i>k'ídahlgwáay</i>	<i>k'i_hihldayáa</i>	<i>tlaadáa</i>
967 1144 senator	<i>lableedgáa</i>	<i>dáalaa'_la'áaygaa</i>	<i>'láanaa'_la'áaygaa</i>	<i>ja_kíl_gudáng</i>	<i>kihl_gáajuwa</i>	<i>xid</i>	<i>táwk'aa'_la'áaygaa</i>	<i>git'iyáa</i>	<i>gaayhldáa'_la'áaygaa</i>	<i>súljuusgaa</i>
977 1155 Senate	<i>sgalagúng</i>	<i>sgawíi</i>	<i>kúnadiiyaa</i>	<i>k'awad</i>	<i>ja_kíl_gudáng</i>	<i>gu_hlgá'áangwaa</i>	<i>sgaadiisdlíyáa</i>	<i>gangáa</i>	<i>kihl_gáajuwa</i>	<i>sgaawnáng</i>
978 1156 violence	<i>kihl_kúunangaa</i>	<i>tláng_ánts'ayaa</i>	<i>'láanuu_sk'iila</i>	<i>is_ánts'ayaa</i>	<i>jáaguugaa</i>	<i>k'ahlgad</i>	<i>kayúugaa</i>	<i>gáwögahl</i>	<i>xa_tláagang</i>	<i>gyáa_k'wíi</i>
991 1171 civil	<i>ja_dáang</i>	<i>dáng_k'íjuu</i>	<i>dáang</i>	<i>ja_dáanggaa</i>	<i>ja_kíl_gudáng</i>	<i>kihl_yaagahl</i>	<i>kayúugaa</i>	<i>íinaa_sáldayaa</i>	<i>gagad</i>	<i>siik'aadiyaa</i>
1009 1190 institution	<i>dáalaa'_la'áaygaa</i>	<i>gyáa_sahl_náa</i>	<i>kehsgadáa</i>	<i>ja_kíl_gudáng</i>	<i>k'úudangaa</i>	<i>tláasala</i>	<i>hlk'ámala</i>	<i>gu_ga_k'áawgaa</i>	<i>k'íijuwaa</i>	<i>xángyaang</i>
1666 1872 review	<i>kíngusganang</i>	<i>keewsganáng</i>	<i>kíndä_gad</i>	<i>kínggwang</i>	<i>gu_kúnjuu</i>	<i>k'áaw_kíng</i>	<i>kíl_xángisälang</i>	<i>kíng_gudáng</i>	<i>hlgwäyáa</i>	<i>gyáajuug_súu</i>
2442 2674 professional	<i>tláajuu'_láa</i>	<i>kíl_míits'angaa</i>	<i>gu_míits'angaa</i>	<i>tlakmíits'angaa</i>	<i>kíl_dlaagáa</i>	<i>kehsgadáa</i>	<i>kíl_k'adangáa</i>	<i>gu_gat'agangáa</i>	<i>gaagúng</i>	<i>k'adangáa</i>
4636 4925 campaign	<i>kayúugaa</i>	<i>kihlgáng</i>	<i>t'ánskad_skúnagalaa</i>	<i>gín_sdaláang</i>	<i>xu_hlkílgangaang</i>	<i>sgawíi</i>	<i>kínyaaw</i>	<i>gwáa</i>	<i>'láanaa'_la'áaygaa</i>	<i>ka_gudunáng</i>

# Results – qualitative evaluation

- high-quality result
  - top search result for a missing word is either synonymous with, or highly semantically related to, the query word in question
    - 'policy' → *wiyasiwêwin* ("law, rule, decision, council, band council, office")
    - 'attorney' → *oyasiwêwiyiniw* ("band councillor, court judge, lawyer")
    - 'administration' → *okimâwiwin* ("chieftancy, social status, government")
    - 'pdf' → *masinahikan* ("book, letter, mail, written document, ...")
- 18/26 – Plains Cree
- 18/26 – Arapaho
- 5/26 – Northern Haida

# Results – qualitative evaluation

- moderate-quality result
  - top match is broadly, but not precisely, semantically related to the query word;
    - 'international → *opîtatowêw* ("Ukrainian, European")
      - a semantically related, but decidedly distinct and non-synonymous term
  - 3/26 – Plains Cree
  - 3/26 – Arapaho
  - 11/26 – Northern Haida

# Results – qualitative evaluation

- poor result
  - top match is either entirely semantically unrelated, or sufficiently irrelevant to be of no use
    - 'percent' → *nisto-sôniyâs* ("three quarters, seventy-five cents")
    - 'career' → *ispîhtaskîwin* ("season")
  - 5/26 – Plains Cree
  - 5/26 – Arapaho
  - 10/26 – Northern Haida

Language	SearchTerm	Headword	CVD	Quality	Definition
Cree	percent	<i>nisto-sôniyâs</i>	0.4470613599	P	three quarters, seventy-five cents
Cree	percent	<i>nîso-sôniyâs</i>	0.466712296	P	two quarters, fifty cents
Cree	percent	<i>pêyak-sôniyâs</i>	0.477778852	P	one quarter, twenty-five cents
Cree	percent	<i>mitâtahto-pîwâpiskos</i>	0.4820271134	P	ten cents
Cree	percent				poorly done, hurriedly done, low quality work, quickly, roughly, without care, sparingly, delicately
Arapaho	percent	<i>nónou3ó.o</i>	0.4391120672	P	million, countless, endless, infinite
Arapaho	percent				quarter (coin), quarterback, in football, quarter (fraction), step- or half- (relative)
Arapaho	percent				grow better, better than average or than before, grow more, more than average or more than before, grow past or beyond a certain point
Arapaho	percent	<i>cébe'eini.ii'óó-</i>	0.5305850506	P	
Arapaho	percent	<i>bébenéh</i>	0.5589234233	P	approximately, about, about, approximately, roughly
Arapaho	percent	<b>koh'u-</b>	<b>0.5904055238</b>	M	half
Haida	percent	<i>kwáadägaa</i>	0.5391592383	P	for S to cost twenty-five cents, cost a quarter
Haida	percent	<i>bíid_inuyáa</i>	0.5666357577	P	for S to cost five cents; to be worth five cents
Haida	percent	<i>ka_dluwíi</i>	0.6392741799	P	for S to decrease in size a bit from previous years [said of salmon]
Haida	percent	<i>t'ánsгад_hlgahl</i>	0.6478499472	P	for S to shrink C in the wash [said of clothes only]
Haida	percent	<i>ináasdlayaa</i>	0.6545500457	P	for S to have grown, have increased in size

Cree	national	pîkaniwiyinînâhk	0.5617200136	H	Piegan country, in the Piegan country
Cree	national	nâtakâm	0.5834629536	P	across the water, across the water, in the bush country, in the north, to the north, Towards land (), Shore.()
Cree	national	pêyakôskân	0.5854004025	H	one family, one tribe, one nation
Cree	national	okimânâhk	0.5908145607	H	government, federal or provincial government, Band Council, Band authorities
Cree	national	nâtakâsi-	0.5924300253	P	into the bush country, wading to shore
Arapaho	national	ciiskóhu-	0.6113572121	P	drive a long ways, run a long ways, run cross-country, do cross-country (the sport)
Arapaho	national	wóóhonéníteeno'	0.6266040206	H	United Nations (the global organization)
Arapaho	national	wóóhoncé'eséítiihího'	0.6266040206	H	United Nations (the global organization)
Arapaho	national	wóóhonítee-	0.6420043409	M	joined as one, in a unity
Arapaho	national	céeciisikoohúútooni'	0.6457337737	P	cross-country running, the sport of cross-country
Haida	national	nánt'as	0.6768269539	P	for S to continue, go on, keep on [said of a story, road, day, etc. or a nominal referring to an event or state]
Haida	national	gyáa_sahl_náa	0.6994960606	P	for S to inherit someone's social position after that person's death
Haida	national	k'án_gudáng	0.7059308887	P	for S to hope to regain their health
Haida	national	ga_gasgadáa	0.710419625	P	for S to have become stirred up [said of a group of people or a town]
Haida	national	chalagúng	0.7137462795	P	for S (pl) {cha} to fall through the air

Cree	author				
Cree	author	<b>masinahikan</b>	0.459487021	H	book, letter, mail, written document, report, paper, magazine, will, A book, (), A letter, (), Paper (), Magazine. ()
Cree	author	<b>nîkânitwêwin</b>	0.4895590544	H	preface, foreword (of a book)
Cree	author	kihci-masinahikan	0.569383353	M	Bible, good book, important book, The Bible. ()
Cree	author	nêhiyawasinhikan	0.5842410922	M	Cree book, Cree bible
Cree	author	nêhiyawi-masinahikan	0.5842410922	M	Cree book
Arapaho	author	<b>hoo3íte.íhii</b>	0.5516592264	H	narrator, storyteller
Arapaho	author	<b>beise'íyei.it</b>	0.5632634461	P	scientist
Arapaho	author	<b>beehíiníisitiit</b>	0.5799974203	P	Creator
Arapaho	author	wo3ónohéíhiinóó-	0.589972645	M	written
Arapaho	author	wo3onohu-	0.589972645	M	written, printed, drawn, decorated, painted, written
Haida	author	<b>ta_k'áalaang_’la’áaygaa</b>	0.7519901246	H	for S to be a writer, secretary
Haida	author	<b>ta_k'áalang</b>	0.7873809636	H	for S to write letter(s), do some writing
Haida	author	st'íi_kéeng_’la’áaygaa	0.793408975	M	for S to be a doctor, physician
Haida	author	sgalangáá	0.7974267751	M	for S to be a composer
Haida	author	<b>táwk’aa_’la’áaygaa</b>	0.8096348643	P	for S to be a gardener, farmer

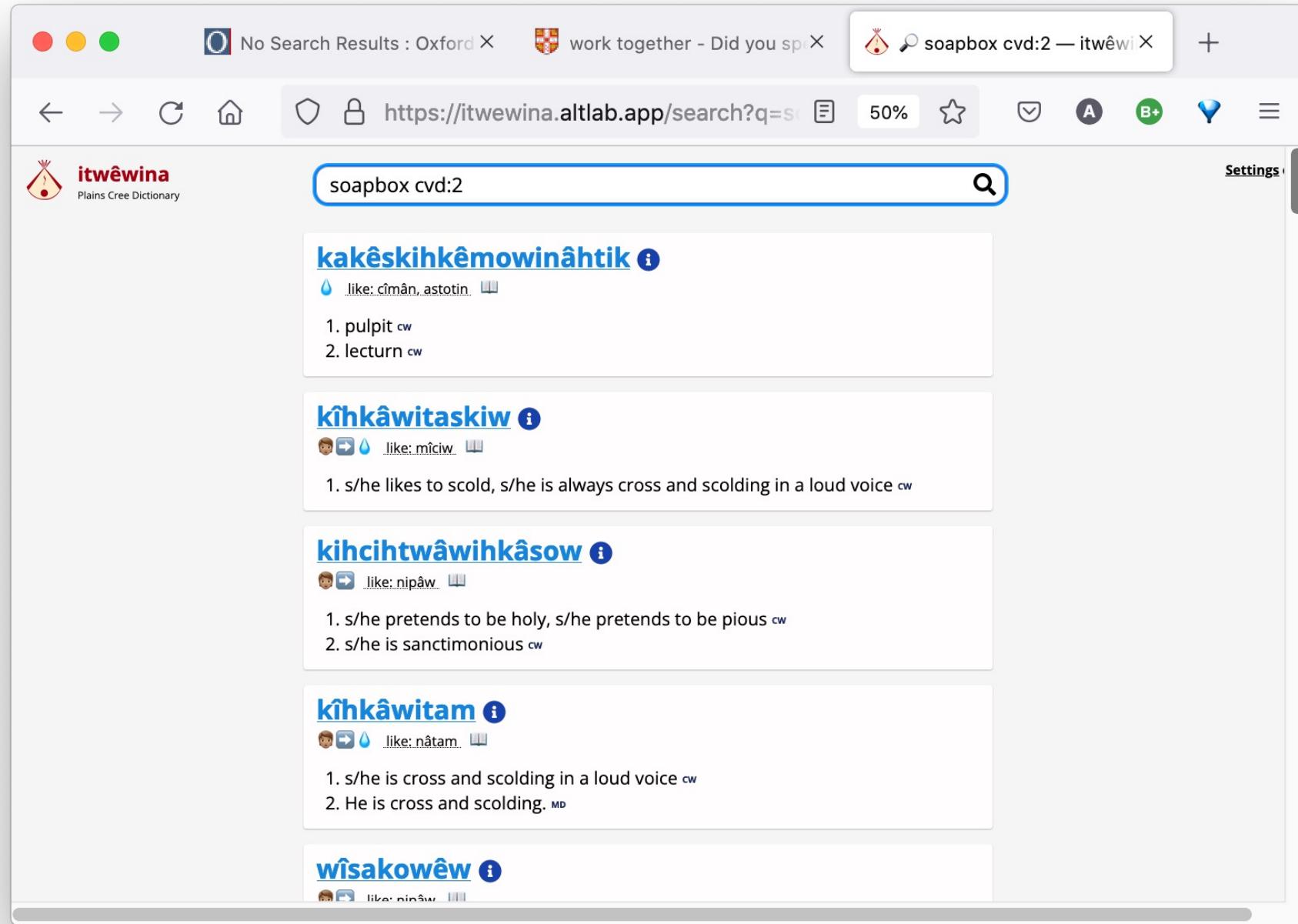
Mean counts of qualitative classes among top-10 ranked matches for 26 “missing” words

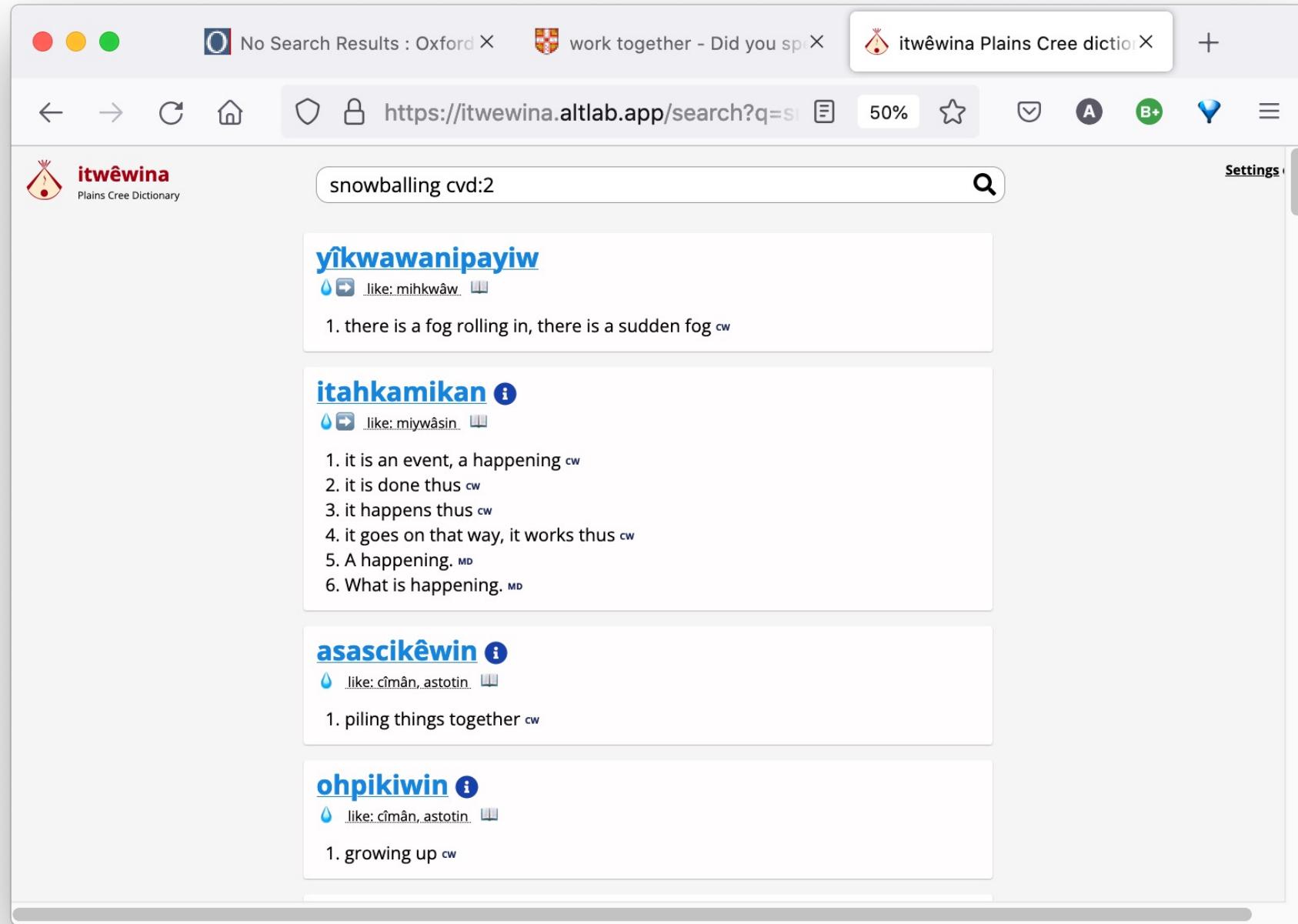
	High	Moderate	Poor
Plains Cree	3.08	3.23	3.08
Arapaho	3.65	3.15	3.19
Northern Haida	0.54	3.23	6.23

Median rank position of semantically most relevant result among top-10 candidates

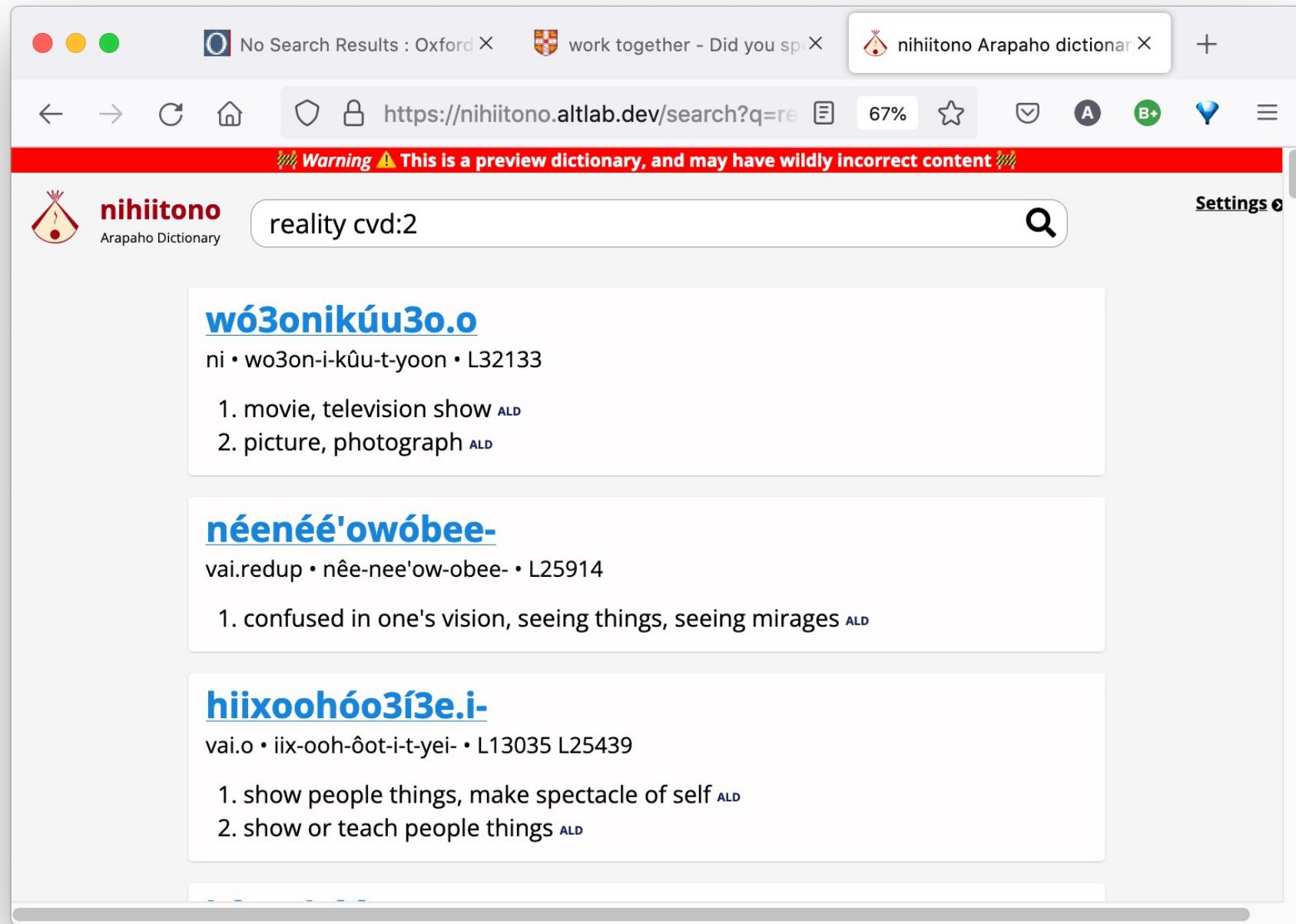
- #2 for Plains Cree
- #3 for Arapaho, and
- #4.5 for Haida

# Metaphorical meanings

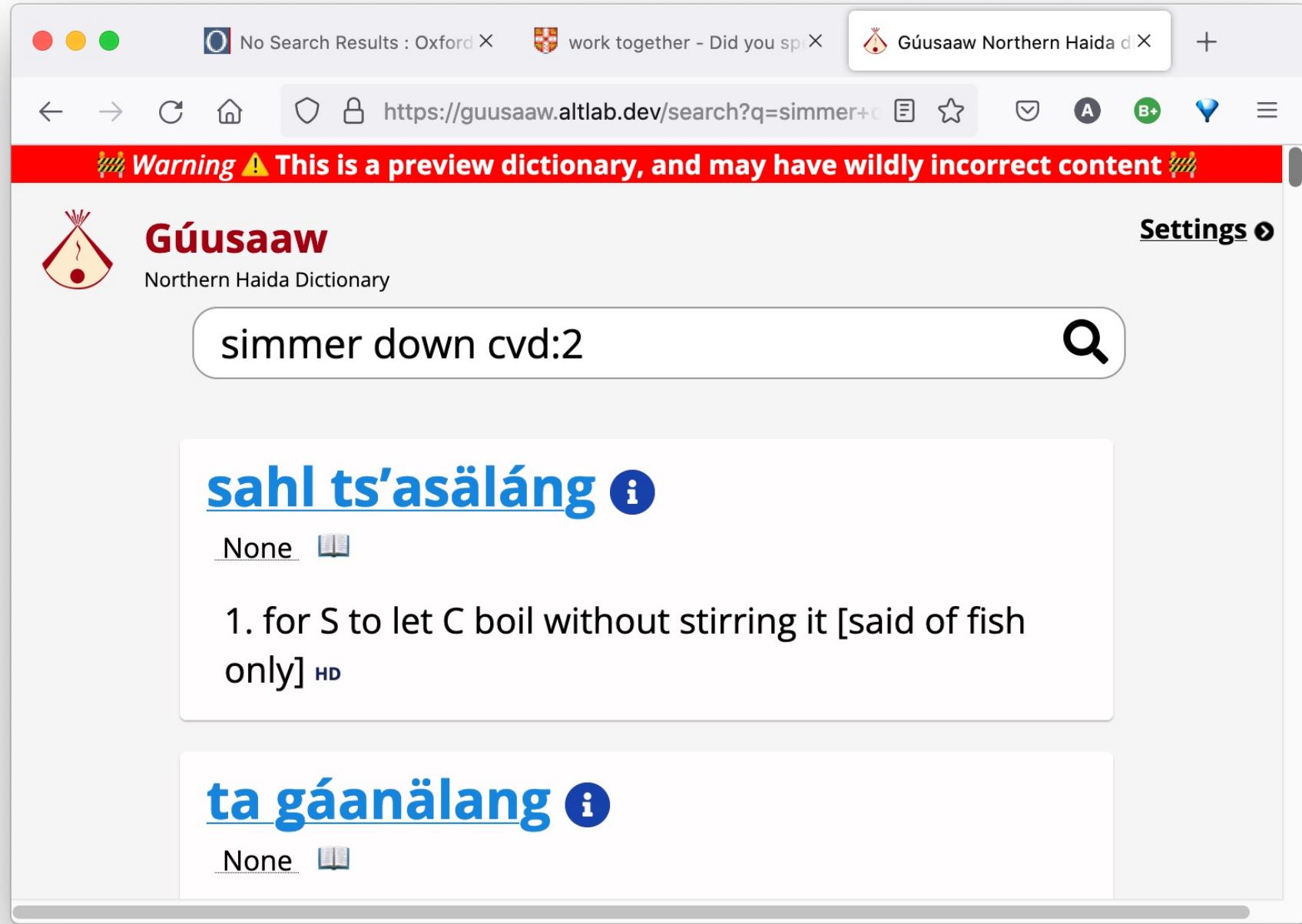




(Goodness of) results depends on contexts  
evident in the corpus



# Multi-word expressions



No Search Results : Oxford X work together - Did you sp X see eye to eye cvd:2 — iX +

← → C ⌛ 🔒 https://itwewina.altlab.app/search?q=see+eye+to+eye+cvd:2 50% ⭐ ⓘ A B+ ⚡ ⏺

 **itwêwina**  
Plains Cree Dictionary

see eye to eye cvd:2

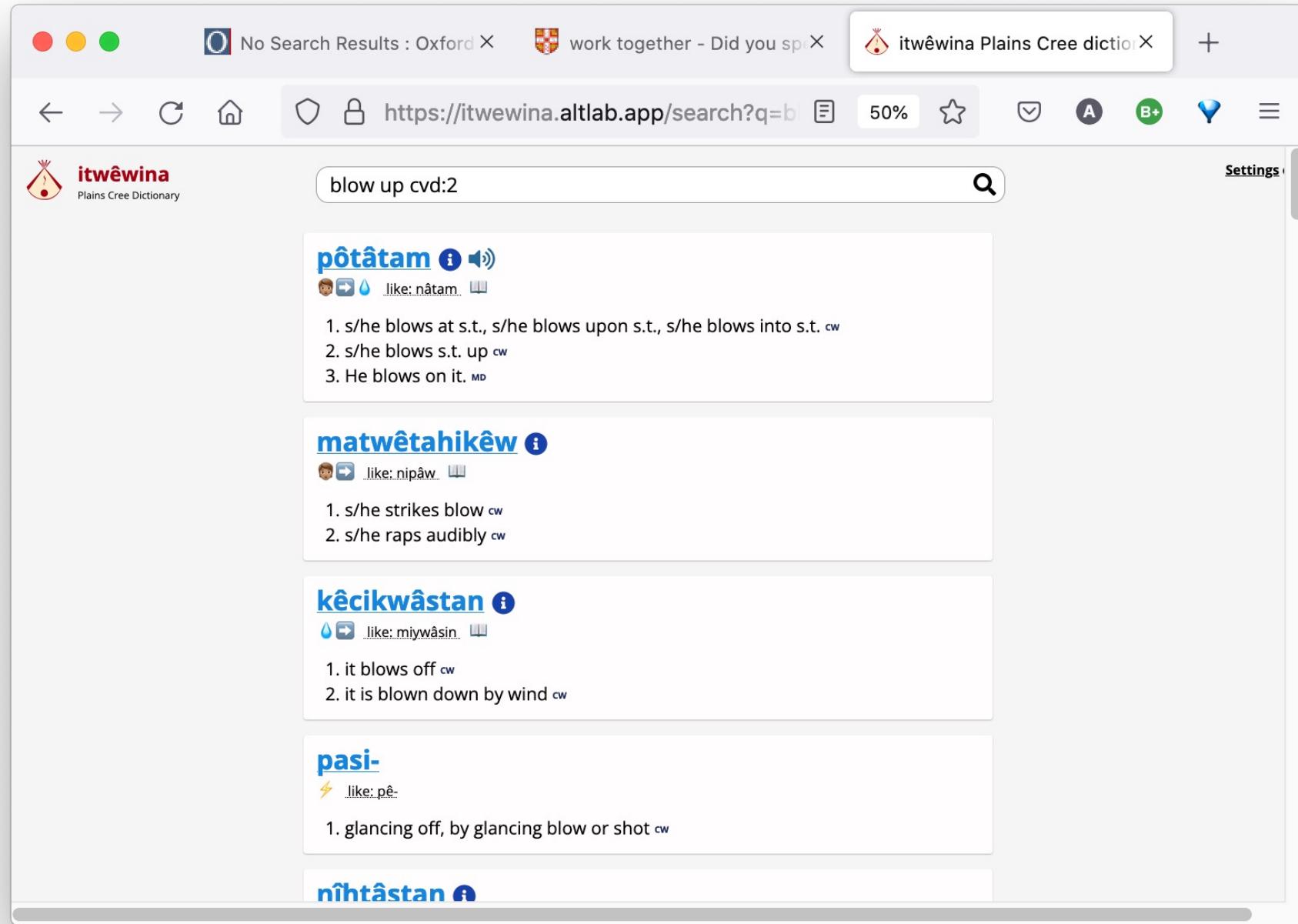
**miskîsikos**  
None like: mitêh (-têh-), mîpit (-îpit-) ⓘ  
1. eye cw  
2. small eye, little eye cw

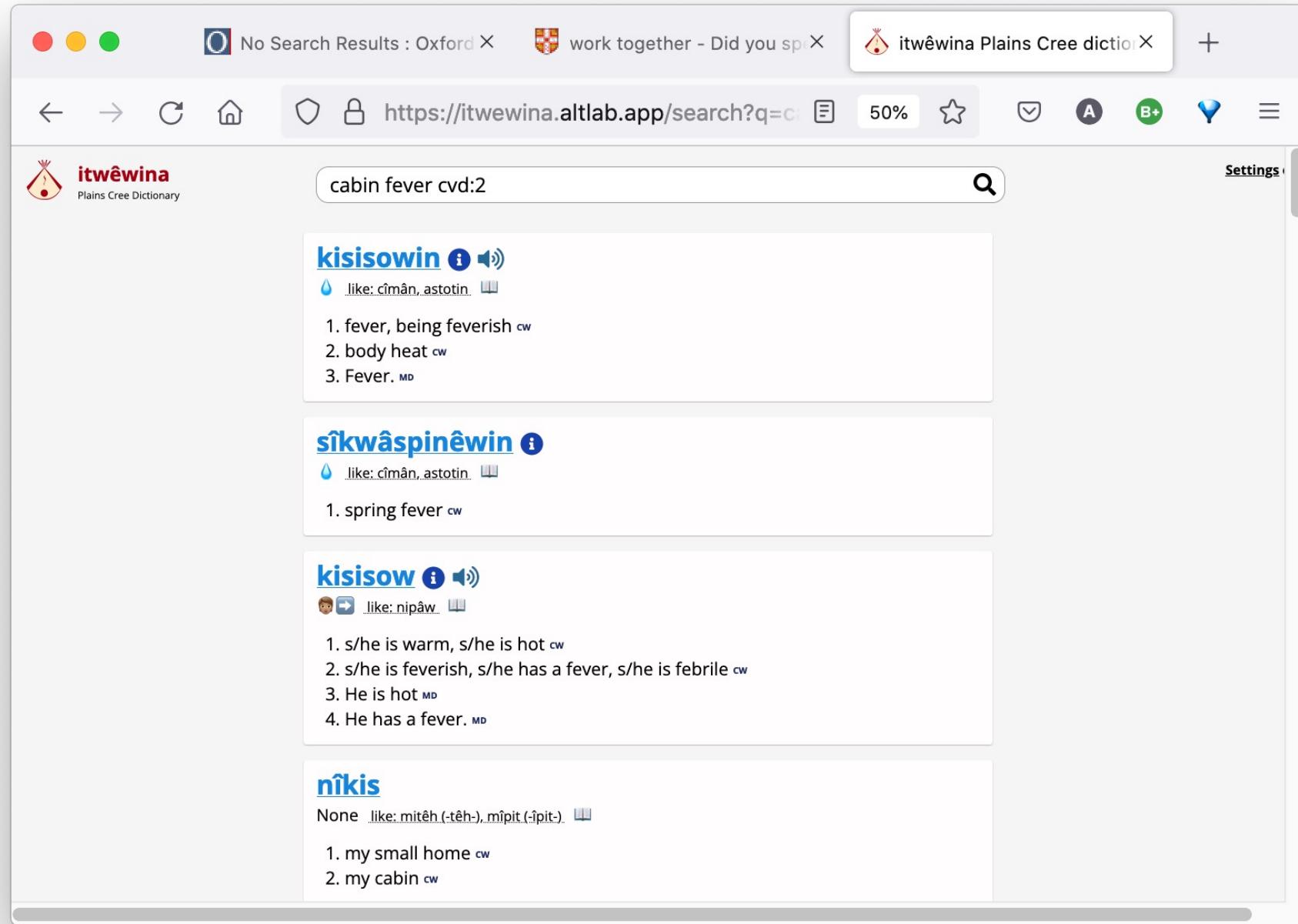
**miskîsik** ⓘ  
None like: miskîsik, mayakask ⓘ  
1. eye cw

**niskîsik** ⓘ  
None like: miskîsik, mayakask ⓘ  
1. my eye cw

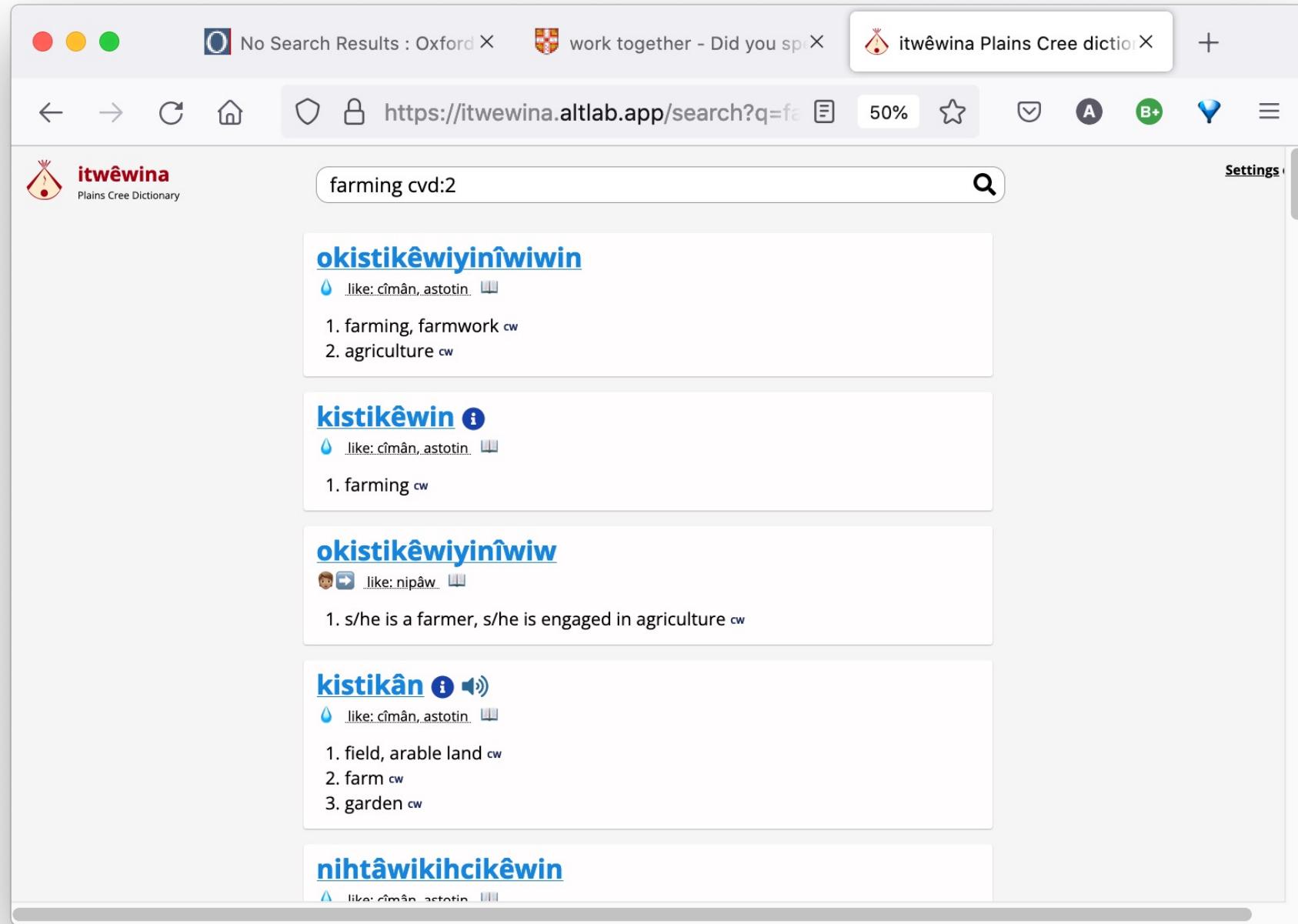
**napatêyâpiw** ⓘ  
like: nipâw ⓘ  
1. s/he looks with one eye cw  
2. s/he has only one eye cw

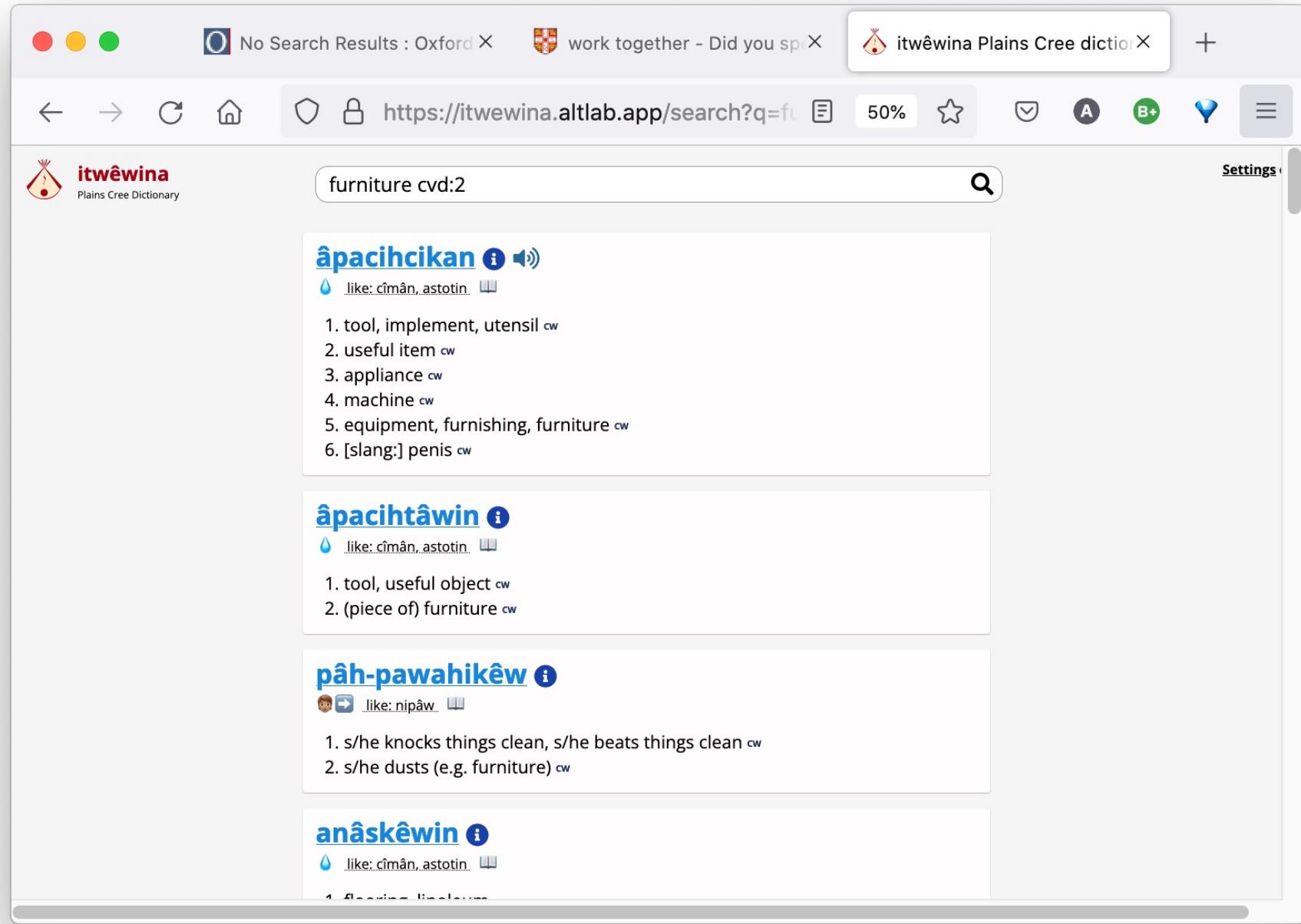
**niskîsikos**  
None like: mitêh (-têh-), mîpit (-îpit-) ⓘ





# Exploring semantic domains







No Search Results : Oxford X



work together - Did you sp X



itwêwina Plains Cree dictio X



https://itwewina.altlab.app/search?q=fu



50%



## anâskêwin ⓘ

💧 like: cîmân, astotin ⓘ

1. flooring, linoleum cw
2. something used as matting cw

## oyâkanikamikohkêwin

💧 like: cîmân, astotin ⓘ

1. cabinetwork, making cabinetry cw

## aspisimowin ⓘ

💧 like: cîmân, astotin ⓘ

1. mattress cw

## anâskâsimon ⓘ

💧 like: cîmân, astotin ⓘ

1. mattress cw

## aspisimon ⓘ

💧 like: cîmân, astotin ⓘ

1. mattress cw

## ayiwin ⓘ

💧 ... ⓘ



No Search Results : Oxford X

work together - Did you sp X

itwêwina Plains Cree dictio X

<https://itwewina.altlab.app/search?q=fu>

50%



## ayiwin ⓘ

like: cimân, astotin ⓘ

1. article of clothing cw
2. [plural:] clothing, apparel cw

## ayiwinisêkin ⓘ

like: pahkêkin(w), askêkin(w) ⓘ

1. cloth, cloth for making clothing cw
2. cloth bag cw

## wâskahikanâhtik ⓘ

like: masinahikanâhtik(w), askihk(w) ⓘ

1. house timber cw

## pâhkwastêwinikan ⓘ

like: cimân, astotin ⓘ

1. clothes dryer cw

## anâskêw ⓘ

like: nipâw ⓘ

1. s/he puts flooring down, s/he lays carpet, s/he puts down linoleum cw
2. s/he lays mats, s/he has a mat cw
3. s/he spreads a blanket cw
4. s/he uses s.t. as a covering cw

# Limitations → Further solutions

- if there is no semantically related or close definition at all in the dictionary, the semantic vector-distance ranking cannot fix that
  - Up to the lexicographers
- if the meaning of a multi-word phrase is not compositional, such a meaning will not be matched with semantically similar single word
  - blow up <-> explode, see eye to eye <-> understand
  - sentence-based language model such as BERT (Bidirectional Encoder Representations from Transformers)
- can increase the rate at which words acquire connotations by analogy with English
  - computational similarity ranking no replacement for human understanding of appropriateness

# Conclusion

- Using vector-space distance can turn search results in the majority language of the definitions more relevant and better ranked
  - semantically relevant matches in ways not achievable with classic information retrieval techniques
  - users can perform successful searches for words that do not occur at all in the dictionary
- These techniques are directly applicable to any bilingual dictionary providing translations between a high and low-resource language.