

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

Antti Arppe^a, Eddie Antonio Santos^a, Andrew Neitsch^a, Jolene Poulin^a, Atticus Harrigan^a, Katherine Schmirler^a, Daniel Hieber^a, Arok Wolvengrey^b

University of Alberta^a, First Nations University^b

We will present the latest developments of *itwêwina* (URL: <https://itwewina.altlab.app/>), a morphologically intelligent version of an on-line dictionary for Plains Cree. This is based on the integration of a computational model of the complex word structure of Plains Cree (e.g. Snoek et al. 2014; Harrigan et al. 2017), with the extensive content of a lexical database for Plains Cree (including English glosses) underlying *nêhiyawêwin : itwêwina / Cree : Words* (Wolvengrey 2001) as well as the *Maskwacîs Dictionary of Cree Words / Nehiyaw Pîkiskweninisa*, (Maskwachees Cultural College 2009), currently amounting to over twenty thousand Cree lexical entries.

The computational morphological model allows for the analysis and recognition of practically all inflected forms of the Cree words in the lexical database, linking these inflected forms with all the information provided for each lexical entry, and the dynamic generation of inflectional paradigms for every lexical entry in the database. Moreover, using novel computational transcriptors we can recognize and present all linguistic content with the current common variants of Plains Cree orthography, whether using Standard Roman Orthography with circumflexes or macrons for long vowels, or Cree syllabics. As the newest development, we have developed transcriptors to “inflect” English definition phrases/sentences so that we may present approximate translations corresponding to inflected Cree word-forms, as well as analyze simple English phrases to directly match these with the corresponding inflected Cree word-forms.

Furthermore, we are continuing to incorporate within *itwêwina* the results of a joint project with the Maskwacîs Education Schools Commission (Littlechild et al. 2018) for creating a dictionary of spoken Maskwacîs Cree (based on the *Maskwacîs Dictionary of Cree Words*), to provide examples of the pronunciation of words alone and in utterances. Moreover, we have reconsidered the software code and user interface design to best work on mobile devices and by non-linguistic/non-academic end-users, which is where we foresee such an on-line dictionary to be most used in the future. Finally, we have leveraged recent advances in semantic vector models for English, a well-resourced language that is the target language in *itwêwina* (Harrigan & Arppe, 2021; Dacanay et al. 2021), to improve the relevance of the Cree results, when searching with English words or phrases, as well as provide semantically close results when exact English word matches cannot be found.

Besides presenting the basic functionalities of *itwêwina*, we will discuss design decisions that we have made, and changed, over time, in reaction to feedback from pilot users in the academia and Cree communities.

REFERENCES

Harrigan, Atticus G., Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud & Arok Wolvengrey (2017). Learning from the computational modelling of Plains Cree verbs. *Morphology*, 27(4), 565–598.

Harrigan, Atticus & Antti Arppe (2021). Leveraging English Word Embeddings for SemiAutomatic Semantic Classification in Nêhiyawêwin (Plains Cree). *Proceedings of Americas-NLP*, June 2021.

Maskwachees Cultural College (2009). *Maskwacîs Dictionary of Cree Words / Nehiyaw Pîkiskweninisa*, Maskwacîs.

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, Miriam Buffalo, Debora Young, Ivy Raine, Paula Mackinaw, Norma Linda Saddleback, Renee Makinaw, Atticus Harrigan, Katherine Schmirler, Dustin Bowers, Megan Bontogon, Sarah Giesbrecht, Patricia Johnson, Timothy Mills, Jordan Lachler & Antti Arppe (2018). Towards a spoken dictionary of Maskwacîs Cree, *Stabilizing Indigenous Languages Symposium (SILS)*, University of Lethbridge, 9 June 2018.

Dacanay, Daniel, Atticus Harrigan & Antti Arppe (2021). Computational Analysis versus Human Intuition: A Critical Comparison of Vector Semantics with Manual Semantic Classification in the Context of Plains Cree. *Proceedings of ComputEL-4*, March 2-3, 2021.

Snoek, Conor, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen & Trond Trosterud (2014). Modeling the Noun Morphology of Plains Cree. In: *Proceedings of ComputEL: Workshop on the use of computational methods in the study of endangered languages*, 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, 26 June 2014, 34-42. ACL Anthology.

Wolvengrey, Arok (2001). *nêhiyawêwin: itwêwina / Cree Words*. Regina: Canadian Plains Research Center.