



**UNIVERSITY OF ALBERTA**

**Evaluating *nêhiyawêtan*:  
A computer assisted language learning (CALL) application for Plains Cree**

**BY**

**Megan Bontogon**

**A THESIS  
SUBMITTED TO THE FACULTY OF ARTS  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
BACHELOR OF ARTS**

**DEPARTMENT OF LINGUISTICS**

**EDMONTON, ALBERTA**

**January, 2016**

UNIVERSITY OF ALBERTA  
FACULTY OF ARTS

The undersigned certify that they have read and recommend to the Faculty of Arts for acceptance, a thesis entitled “Evaluating *nêhiyawêtan*: A computer assisted language learning (CALL) application for Plains Cree,” submitted by Megan Bontogon in partial fulfillment of the requirements for the degree of Bachelor of Arts.

.....  
Honours Thesis Supervisor

.....  
For the Department

## ABSTRACT

*nêhiyawêtan*, a computer assisted language learning (CALL) application for Plains Cree was evaluated. User reactions and behaviours were analysed using heuristic and co-operative evaluation methods. Feedback of heritage learners and other university students on the user interface were compared. This feedback helped to create an updated version of *nêhiyawêtan* by adding features that allowed for more diverse user strategy to allow the program to assist learners in a university or community setting. The findings serve as a guideline for evaluating and creating CALL programs for other Canadian indigenous languages (CILs).

## PREFACE

The human research ethics approval necessary to conduct the research described in this work has been obtained from the Research Ethics office of the University of Alberta. Approval for the user evaluations covered in Chapter IV can be found under the study ID: MS8\_Pro00023436, while the amendment approval needed to extend to this work can be found under amendment/renewal ID: Pro00023436\_AME5.

The author created Tables 1.1-2 which have been modified from Table 1 of Marian et al. (2007). Additionally the author constructed Figure 2.1, Tables 2.1-2.8, B1.1, and the severity ratings on page 56 which have been adapted from Hagdahl (2002) Figure 6, Tables 5-11, Table 2 and the severity ratings on page 83, respectively. The author was the lead investigator and was responsible for concept formation, conducting user evaluations, data collection and analysis, as well as manuscript composition.

*nêhiyawêôtân* itself was made in collaboration with the ALTlab at the University of Alberta, and Giellatekno at UiT the Arctic University of Norway. The initial demonstration version of *nêhiyawêôtân* was created by Lene Antonsen, M.A., which was then expanded to the one evaluated in this thesis through contributions by Dorothy Thunder, M.A., Antti Arppe, M. Sc., Ph.D, Ryan Johnson, M.A., Atticus Harrigan B.A., and the author. Lene Antonsen, M.A., and Ryan Johnson, M.A., played major roles in ensuring the application was in appropriate condition for evaluating. The Faculty of Native Studies and Dorothy Thunder, M.A., supplied the course materials used in the program. Antti Arppe, M.Sc., Ph.D., and Jordan Lachler, Ph.D., were supervisory authors on this project, and were involved throughout the project in concept formation and manuscript edits.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor Antti Arppe, M.Sc., Ph.D., at the University of Alberta, who supported me during the project and gave me valuable feedback and references. Because of his support I was able to participate in meetings with scholars specializing in Cree from across Canada, and also meet with scholars from Europe specializing in computer assisted language learning (CALL) and computational modeling on multiple occasions. I was also given the opportunity to undertake field work in Maskwacis, A.B. which has been a great experience. I very much appreciate all the opportunities I have been given in relation to this project.

I would also like to thank Jordan Lachler, Ph.D., for being my second reader and providing me with valuable input on the study. As well as providing me with the opportunity to meet with Marianne Ignace, Ph.D., and others working on a collaborative CALL project at Simon Fraser University and discuss the state of other CALL programs for Canadian Indigenous Languages. I am truly thankful for this opportunity.

I also wish to acknowledge Lene Antonsen, M.A., for her support and input into the evaluation design process. She and Ryan Johnson, M.A., also played a crucial role in ensuring that the demonstration version was ready for testing and their help has been greatly appreciated.

Assistance provided by Herb Colston, Ph.D., for helping with editing and writing of this thesis has been very much appreciated.

Thanking Dorothy Thunder, M.A., and the Faculty of Native Studies (NS) for providing student participants and electronic access to the NS teaching materials, respectively.

A special thanks to Benjamin V. Tucker, Ph.D., for allowing me the use of the Phonetics Laboratory space in order to run this study.

## TABLE OF CONTENTS

CHAPTER I. Introduction .....	1
1.1 CALL for CILs .....	1
CHAPTER II. Literature Review .....	4
2.1 Previous Evaluation Methods .....	4
2.2 Evaluating CALL Feedback and Functions .....	4
CHAPTER III. Background .....	6
3.1 Origins of <i>nêhiyawêtan</i> .....	6
3.2 <i>nêhiyawêtan</i> .....	6
3.3 CALL Application Comparison .....	11
CHAPTER IV. Method .....	14
4.1 Checklist .....	14
4.2 Participants .....	14
4.3 Study Outline .....	17
CHAPTER V. Results .....	21
5.1 Navigation .....	22
5.2 General Interface Design & Template Structure .....	22
5.3 Observing Use of Functions .....	22
5.4 Evaluating Verb Exercise Templates .....	24
5.5 Error Analysis .....	24
5.6 User Performance .....	25
5.7 ‘Show Answers’ .....	26
5.8 ‘Help’ Feedback .....	27
5.9 Self-Correction .....	28
CHAPTER VI. Discussion .....	29
6.1 Navigation .....	29
6.2 ‘Click on a word’ .....	29
6.3 Instructions .....	29
6.4 The ‘help’ Function .....	30
6.5 ‘Test Answers’ and ‘Show Answers’ .....	30
6.6 ‘New Set’ .....	32
6.7 Audio .....	32
6.8 Content Relevance .....	32
6.9 Question Formatting and Interface Options .....	33
CHAPTER VII. Conclusion .....	35
REFERENCES .....	37
APPENDIX A: A Design Study .....	40
APPENDIX B: Study Checklist .....	45
APPENDIX C: User Evaluation .....	49
APPENDIX D: Suggested Improvements & Heuristic Evaluation .....	52

## LIST OF TABLES AND FIGURES

Figure 1.1 .....	7
Figure 1.2 .....	8
Figure 1.3 .....	8
Figure 1.4 .....	9
Figure 1.5 .....	10
Figure 1.6 .....	10
Table 1.1 .....	15
Table 1.2 .....	16
Figure 2.1 .....	17
Figure 2.2 .....	19
Table 2.1 .....	21
Table 2.2 .....	22
Table 2.3.....	23
Table 2.4 .....	23
Table 2.5 .....	23
Table 2.6 .....	23
Table 2.7 .....	24
Table 2.8 .....	24
Figure 3.1 .....	25
Table 3.1 .....	26
Table 3.2 .....	27
Table 3.3 .....	28
Table 3.4.....	28
Figure 4.1 .....	31
Figure 4.2 .....	31
Figure 4.3 .....	33
Figure 4.4 .....	33
Table A1.1 .....	41
Table A1.2 .....	42
Table A1.3 .....	42
Table A1.4 .....	43
Table B1.1 .....	46

## CHAPTER I. INTRODUCTION

Computer Assisted Language Learning (CALL) research, has largely ignored CALL in endangered languages; even rarer are studies addressing CALL in endangered Canadian Indigenous languages (CILs). This is not surprising, given the oral history of many CILs. Rapid urbanization and assimilation of CIL communities has facilitated the loss of their language and culture (Struthers & Peden-McAlpine 2005:1273). This provides a convincing argument for written, audio and video documentation (Struthers & Peden-McAlpine 2005:1273).

Many communities now support these forms of documentation to assist them with intergenerational transmission of their languages. This knowledge can then reach isolated speakers through use of online platforms. Extending the domains of language use in a technology-dominated society is essential when combatting language loss. This introduction will discuss the unique conventions and difficulties faced when creating a CALL application for both Plains Cree and CILs in general.

### *1.1 CALL for CILs*

While CALL using online platforms provides many benefits, it yields initial challenges in both constructing and evaluating new applications. This case study will address these challenges in a detailed user evaluation of *nêhiyawêtân*, a CALL application for Plains Cree. *nêhiyawêtân* (*Let's speak Cree*) is a collaborative project between the Alberta Language Technology Lab (ALTLab) at the University of Alberta, and Giellatekno at UiT the Arctic University of Norway. The general design and interface is based off of *Oahpa!* (*Learn!*), an existing CALL application for the Northern Saami language, and is currently in its demonstration form.

Since the interface used by *nêhiyawêtân* has been removed from its original Norwegian and North Saami context, there have been some complications since its inception. Limited technical knowledge and support, limited time, lack of language consultants, lack of standardized use of writing system, and social constraints all hinder the development of CALL applications for endangered languages (Ward & Genabith 2003:237). These issues have undoubtedly affected the creation of *nêhiyawêtân* in varying degrees as well.

Since the creators of the CALL interface reside in Norway, there is restricted technical support available for the application. The ultimate goal of the design is for it to be accessible in multiple modalities. Although achievable, advanced technical knowledge and hardcoding is still required. This would pose a larger problem for CIL communities with few technical resources. While the ALTLab is able to create the exercise template source files needed to create the application, the transformation of these files into the CALL application still requires the assistance of collaborators at UiT or a collaborator residing in the United States. This inevitably slows the process of design and for many endangered languages this loss of time cannot be afforded. Much of this time loss is attributed to multiple ongoing projects and other commitments of all collaborators, rather our more distant collaborators have been most willing to complete tasks, other constraints permitting.

In addition to constrained technical support, Ward & Genabith (2003:237) mentioned issues due to lack of standardized orthography. Orthographical variance affects both *Oahpa!* and *nêhiyawêtân*. This first instance of *nêhiyawêtân* is meant to compliment the University of Alberta's Native Studies (NS) 152 Introductory Cree course materials. However, some exercises use

orthographic representations that correspond to Wolvengrey (2001). For example, in Plains Cree /e/ is exclusively long. The NS 152 materials represent it as “e”, Wolvengrey (2001) uses “ē”, and *nêhiyawêtân* utilizes “ê.” This connection to Wolvengrey’s material is natural since Wolvengrey (2001) implements the standard roman orthography for Plains Cree and is consistent with an ongoing collaborative ALTLab online dictionary project. The project is an online version of Wolvengrey (2001) and is linked to *nêhiyawêtân* as an additional resource for users.

Vowel length markers can be confusing for beginners and the discrepancy, both between sources and within *nêhiyawêtân* itself, could cause additional problems. *nêhiyawêtân* only accepts answers using “ê.” To resolve issues regarding variance in the orthographic convention and dialectal differences of certain words, such as “*sôniyâwikamik*” vs. “*sôniyâw’kamik*,” *nêhiyawêtân* accepts both answers. However, there are still some other discrepancies between meaning and representation that need to be addressed; such as, “*minihkwâcikan*” (in Wolvengrey 2001 as ‘cup’) vs. “*minihkwêwiyâkanis*” (in Wolvengrey 2001 as ‘glass’, but in NS 152 materials as ‘cup’).

In order to address the problem of social constraints (Ward & Genabith 2003:237) this study will involve a user evaluation of *nêhiyawêtân*. For example, the exercises in the demonstration version have retained their original *Oahpa!* names. These have not yet been changed to culturally relevant names because there are no obvious Cree equivalents. By evaluating the program in its demonstration stage, problems such as these (pertaining to cultural relevance) can be solved early on. Nesbitt (2013) concluded that mid-design user input could help improve application versatility to meet a greater variety of student needs. This is important for the development of *nêhiyawêtân* because although it was created to be used in an academic context, many beginner heritage speakers, and even some fluent speakers enroll in NS 152.

Ideally, *nêhiyawêtân* will function as a culturally relevant academic CALL program and eventually an accessible tool for creating literacy amongst fluent speakers. When communities lack a variety of resources for language learning, it forces one tool to be multipurpose. Junker & Torkornoo (2012) speak to the importance of adapting to the needs of communities when creating online tools with limited resources. Their project has also worked collaboratively to create language games for East Cree and Innu. Although these language games have also been linked to other resources such as online dictionaries (not unlike *nêhiyawêtân*) they still focus mostly on vocabulary, using audio and images to familiarize fluent speakers with the syllabic orthography.

Social constraints also have an effect on the semantic categories that may be discussed and included in the exercises. For *nêhiyawêtân*, this is not as big of an issue since it is based on pre-approved course material. Another related issue encountered while creating exercises is the degree of precision used to create the semantic categories which within questions are grouped. As the ALTLab has used the SIL Rapid Words model to categorize words for elicitation purposes in the past, it was the first to be considered for implementation.

When categorizing these words using the broadest Rapid Words categories, the contents of the sub-categories seems unrelated. For example, when using “6.0 Work and Occupation” as a semantic category, “fish” and “dog” will be generated since they are part of the sub-categories “6.4 Hunt and Fish” and “6.3.1 Domesticated Animal” (which is a sub-category to “6.3 Animal Husbandry”), respectively. When the user selects “Work and Occupation” and is presented with “dog” and “fish” they will understandably be confused. If the sub-categories were to be used, there

would not be a sufficient amount of words to provide acceptable randomization within each question category since *nêhiyawêtan* currently has only 446 words. Additionally, the use of sub-categories would result in an unmanageable amount of semantic categories to choose from. Therefore the decision was made to implement the relevant approximate semantic categories featured in the Native Studies course materials.

## CHAPTER II. LITERATURE REVIEW

This literature review will attempt to highlight various methods for evaluating CALL applications and how they apply or do not apply to this study. Functions of existing CALL programs will also be explained. The strengths and weaknesses of these functions will be weighed.

### **2.1 Previous Evaluation Methods**

In this study, we will attempt to identify user needs through evaluations. This section will address previous evaluation methods. A combination of these methods will be used to evaluate *nêhiyawêtan*.

Antonsen (2013), was able to provide a quantitative analysis of *Oahpa!* question and answer drills. This was done using data collected by the application, which provided an elaborate error analysis. This is not yet possible in the case of the *nêhiyawêtan* demonstration version, as it lacks both the technology and a populated user base. For these reasons, this study takes a more qualitative approach.

Previous studies have compiled user experience surveys using pre- and post-CALL questionnaires, interviews, and observations of both on-task and off-task behaviour (Son 2007). This study will be conducted in a similar manner. The pre-CALL questionnaire will be supplemented by the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al. 2007) and all user activity (or lack thereof) will be observed. There will be no formal interview. Instead the participants will be encouraged to think out loud during their use of the CALL application. This will create an open dialogue between the participant and the researcher, and facilitate co-operative evaluation methods (Monk 1993).

Hagdahl (2002) found this to be an effective method for his user evaluation study. In order to uncover potential user problems, Hagdahl (2002) conducted a design study that preceded the user evaluation component. The design study for *nêhiyawêtan* was completed previously, and relied on heuristic evaluation (also used by Hagdahl 2002) to identify and rate severity of design problems. More information on the design study and its results can be found in Appendix A. For the current study, user activities will be recorded as they complete exercises within the application. Afterward, they will complete a post-CALL questionnaire evaluating *nêhiyawêtan*.

Overall, conducting these user evaluations will help to determine whether *nêhiyawêtan* meets the needs of learners, and if the students are using it as intended. Moreover, it will shed light on the compatibility of *nêhiyawêtan* and the course it is meant to complement. Uncovering the strengths and weaknesses of the program at various levels of completion, will help to create a more versatile program and broader spectrum of use for students of different backgrounds and competencies.

### **2.2 Evaluating CALL Feedback and Functions**

When evaluating an application, it is important to obtain user perceptions of various functions. This section will discuss various functions present in other existing CALL applications and highlight their strengths and weaknesses.

One of the most important functions (though somewhat rare in applications for endangered languages), is effective answer feedback. In language learning overall, it has been found that metalinguistic feedback is helpful for adult second language learners (Carroll & Swain 1993). More

specifically, CALL applications can be used to raise grammatical awareness through metalinguistic feedback (Nagata & Swisher 1995).

Most endangered language CALL involves showing only whether the answer is correct or not. This is the typical “wrong, try again” model which does not give the user enough metalinguistic information to describe how or why the answer does not match the correct form (Nagata 1993:130). With *nêhiyawêtan*, effort has been made to ensure that users have adequate metalinguistic feedback. However, this is difficult to achieve without having personalized feedback.

There are generally outlines four types of answer feedback (Garrett 1987):

1. “Correct” or “incorrect” feedback.
2. Feedback is generated by comparing the given answer letter-by-letter to the correct answer.
3. Standard feedback is given based on anticipated user errors.
4. Intelligent natural language processing, which analyzes the response and provides individualized feedback.

Currently, the *nêhiyawêtan* demonstration version uses the second and third types of feedback, however, the goal of the future version of the application is individualized feedback. Having a natural language processor would allow for more flexibility with regards to user input, and question variability and complexity. Flexibility is the key to allowing the users to explore the language more thoroughly. This is an important part of second language acquisition (Felix 2003; Chapelle 2001). Though as discussed in this study and Heift (2012), a simple fill-in-the-blank exercise can provide fairly specific and accurate feedback on mistakes.

It is important to consider that without natural language processing, “help” feedback messages based on anticipated user errors in *nêhiyawêtan* can easily become lengthy since feedback is not individualized. Although most students actually use CALL system feedback when available, they are overwhelmed by lengthy feedback (Heift 2001). Due to user error variability, there could be multiple errors within a single answer that need addressing. It simply is not possible to provide error messages for all the errors at the same time (Heift 2003). Heift (2001) suggests a prioritized, sequential revealing of these errors. However, this prompts students to only fix the errors outlined by the system (Heift 2001). This brings up the issue of a detrimental dependence on error feedback.

Cobb & Stevens (1996) note that the students that excessively relied on system help, did not learn as much as those trying to answer questions through their own trial and error. Therefore, the goal of future versions of *nêhiyawêtan* will attempt to provide precise and personalized metalinguistic feedback that still challenges the user to correct and formulate their own answer.

It is hoped that the application will become more integrated and comprehensive in order to support semantics, pragmatics, cultural knowledge, and social abilities. This will be achieved by using a combination of the functions previously discussed.

## CHAPTER III. BACKGROUND

The background will explain where and how the open-source code for *nêhiyawêtân* originated. This section will also explain the many functions embedded in the interface of the application. Additionally, the functionalities of *nêhiyawêtân* will be compared to those of other CALL applications for North American indigenous languages.

### 3.1 Origins of *nêhiyawêtân*

Preceding *nêhiyawêtân*, the first *Oahpa!* project began in 2008 (Antonsen 2013). *Oahpa!* is a computer assisted language learning program for North Saami, an endangered language spoken in Northern Europe (Antonsen 2013). This Uralic language has complex morphology, which most learners (typically Norwegian and Swedish nationals) have not encountered in languages learned previously (Antonsen 2013). The initial goal of the project was to create an interactive application to replace textbooks in a university course which would provide students with immediate feedback on answers outside of class time.

Like its predecessor, *nêhiyawêtân* uses finite state transducers (FST) to create an intelligent computer assisted language learning (ICALL) program that is fully productive. The computational morphological model central to both of these ICALL programs is therefore not limited to a set list of common lexemes, but can produce many different inflected forms (Antonsen et al. 2013).

Arguably the most noticeable difference between the two programs is the answer feedback given to users. *Oahpa!* is currently able to generate individualized feedback, whereas *nêhiyawêtân* approximates possible answers and gives feedback based on potential user errors. Errors in user input are recognized in *Oahpa!* and the program is then able to produce feedback relative to the specific error (Antonsen et al. 2013). This process is improved by logging user data. It is hoped that *nêhiyawêtân* will soon have the capabilities of individualized feedback and tracking user data as well.

The computational model seems to work well for the purpose of both programs, since the goals for *Oahpa!* correspond with those for *nêhiyawêtân*. By basing *nêhiyawêtân* off of existing curriculum, it is hoped that the application will provide students with valuable feedback outside of class time. The *Oahpa!* framework was chosen to create this resource since it has the capacity to cater to the complex morphology of Plains Cree.

### 3.2 *nêhiyawêtân*

As previously mentioned, *nêhiyawêtân* is a collaborative project between the Alberta Language Technology Lab at the University of Alberta, and Giellatekno at the UiT the Arctic University of Norway. This ICALL program for practicing plains Cree is currently in its demonstration form and is likely the first ICALL program for a CIL. There are four main sections, “Leksa”, “Morfa-S”, “Morfa-C”, and “Numra”, which can be seen in Figure 1.1. These sections have retained their original *Oahpa!* names, but will be replaced by Plains Cree names in future versions.

There is a menu on the left side of the interface that allows the user to select the part of speech to be practiced, and features a link to a dictionary as a resource (see Figures 1.1-1.5). The instructions for each exercise are featured to the right of the questions for all exercises (see Figures 1.2-1).

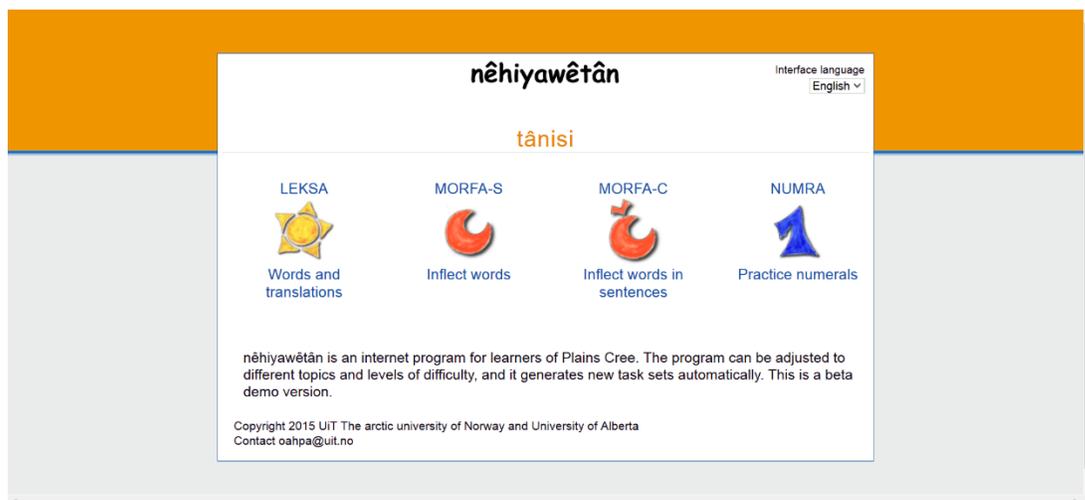


Figure 1.1: The nêhiyawêtan homepage.

“Leksa” lists words (in green for animate, black for inanimate) and the user must give the translation. This exercise can be done from Cree-to-English, English-to-Cree, Cree-to-French, and French-to-Cree. The words can also be sorted by course unit, which corresponds to the materials for NS 152 Introductory Cree at the University of Alberta, or by semantic category (inspired by topics in the NS 152 material). Users have the option to select units in isolation as well as accumulated units.

After completing the questions, the user can submit the answers for testing using the ‘test answers’ function (see Figures 1.2-1.4). Correct answers will become green, and incorrect answers will become red with a red ‘X’ at the end of the text box (see Figures 1.2-1.4). The user then has the option to continue trying to correct their answers, or show the correct answers (see Figure 1.5). Leksa also gives users the option to practice words with audio files (see Figure 1.2), however, the number of these audio prompts is limited. Future versions will show increased numbers of questions featuring audio prompts based on a collaborative dictionary recording project with Miyo Wahkohtowin Education of Maskwacîs, Alberta.

Figure 1.2: *Leksa (Vocabulary Testing)*.

“Morfa-S” is meant for practising producing inflected forms with limited to no context. Within these exercises the user can again select the course unit. These exercises can be made to focus on nouns (PLURAL, DIMINUTIVE, LOCATIVE), verbs (PAST, PRESENT) or possessives (SECOND PERSON SINGULAR, THIRD PERSON SINGULAR). These exercises have the same functions as the last, however, these exercises also feature answer feedback to give the user helpful hints for correcting their answer (Figure 1.3).

Figure 1.3: *Morfa-S. Using the ‘help’ function in a noun inflection exercise.*

In “Morfa-C”, the user can again select from nouns (PLURAL, LOCATIVE, POSSESSIVE) and verbs (PAST, PRESENT). The exercise differs from Morfa-S in that the user practices these paradigms in a discourse context within a question answer setting. These exercises have answer feedback as well (Figure 1.4).

The screenshot shows the MORFA-C interface. On the left is a sidebar with a red crescent moon logo and the text 'MORFA-C'. Below the logo are links for 'Nouns', 'Verbs', and 'Pronouns'. Further down are 'Reference materials' and 'Dictionary'. The main content area has a 'Tense' dropdown set to 'past' and a 'New set' button. There are five questions:

- kiki-nikamon cī otākosihk?  
êha, niki-nikamon otākosihk.
- kiki-kiwânâwâw cī otākosihk?  
êha, kiki-kiwânân otākosihk. ✖ help
- kiki-pimohtân cī otākosihk?  
êha, niki-pimohtân otākosihk.
- kiki-pâhpin cī otākosihk?  
êha, niki-pâhpin otākosihk.
- kiki-otâpâson cī mēskanâhk isi?  
êha, niki-otâpâson mēskanâhk isi.

At the bottom of the main area are buttons for 'Test answers' and 'Show the correct answers', and a score of 'Your score: 4/5'. On the right, a box titled 'Practise past tense' contains the instruction: 'Talk about what people did yesterday. Complete the response to the question by entering the verb in the correct form.' Below that, a hint box states: '"kiwêw" is an animate intransitive (AI) verb and this question is asking for the 1P form. The 1P form of an AI verb uses the ni- prefix and the -nân suffix.'

Figure 1.4: *Morfa-C*. Using the ‘help’ function in an exercise involving verb inflection in the context of a sentence. This function responds to user answers with hints on how to correct any incorrect answers.

“Numra” is a tool for practicing numbers (Figure 1.5). The user can either select “Cardinals”, “Time”, “Date”, or “Money”. For all of these subcategories, the user can select either numerals-to-words or words-to-numerals. When practicing cardinal numbers the user can also choose the number range. “Time” also allows the user to select either the “easy”, “medium” or “hard” difficulty setting. Answer feedback is not currently available for Numra but will be included in future versions.



considered, as all possible sentences within the application (which number in the thousands) cannot be practically recorded.

### ***3.3 CALL Application Comparison***

To supply information on how *nêhiyawêtan* compares to the methods of other CALL programs for indigenous languages, a brief comparison of various applications has been undertaken. For indigenous languages in the context of Canada, there are very few fully developed CALL programs available for learners. The currently existing programs have most of their emphasis on oracy and less opportunity for developing literacy. Many of these programs are also centered on learning vocabulary, and often do not cover sentence structure but merely common basic phrases. To get a full view of what is available for endangered language CALL programs, we will be looking at North America's progress as a whole.

Dakelh, Skidegate Haida, Nisga'a, Secwepemc, Ktunaxa (all of which are indigenous languages of British Columbia) are just some of the CILs that have applications for their languages available for free in iTunes (First Voices 2013). Unfortunately, this makes them only available for Apple products. All of these programs consist of a dictionary and a collection of phrases. This is a big step in expanding the domains these languages are used in, however, more must be done to further increase materials for CALL.

For Android, there are applications for Tlingit, Oneida, Ho Chunk, and Ojibwe. The application for Southern Tlingit has all of the vocabulary listed in English, with audio in Tlingit (Roberts 2015). The Oneida application which features only basic phrases also has English as the dominant interface language for the application (Sovereign Oneida Nation of Wisconsin 2015). The difference, however, is that once selected from the main page, the pages for each individual phrase have the Oneida translation as well as audio files featuring the phrase in Oneida; though the Oneida translation is still featured after the English in a small font. Nitsi Poh Sin, an application for Ojibwe also has all its vocabulary listed in English, with only audio files in Ojibwe (Piikani Societies 2015).

The Ho Chunk application is a bit more complex, but still focused mostly on oracy (Ho Chunk Nation 2015). It still has the main menu in English, but has pictures and audio as well as the vocabulary written in the Ho Chunk orthography. Aside from the pictures, this application also stands out from the rest because it has games and quizzes for the vocabulary as well. After picking a semantic category, one can then select "Learn", "Games" or "Quizzes".

The "Learn" function provides the learner with a list of terms in English. Once selected, a picture is shown and the Ho Chunk audio is played. Underneath the picture in small font it displays the word in Ho Chunk orthography. There is also the option to record yourself saying the terms, and compare your pronunciation to that of the audio file.

The "Games" function lets the user select from easy, medium and hard difficulty levels. The game shows four pictures one at a time, each with its own audio. After all of the pictures have been revealed, one of the terms is repeated. The user must select the picture that represents the audio.

The "Quizzes" function has three subcategories: "Listening", "Speaking", and "Reading." The "Listening" option shows four pictures, and the user must choose which one best represents the audio. This differs from the games in that it does not individually describe each picture first. The "Speaking" option shows the user a picture. The user then records what the picture is using their

microphone, they are then able to compare their recording to the intended recording. The “Reading” option shows a picture and has four Ho Chunk words written beneath it in small print. The user is to pick the word represented by the picture.

For websites, there is Dene K’ę́ Gúdeh, which again has mostly English as the interface language (Dehcho Friendship Center 2015). The site seems as if it is more geared towards children than the applications previously discussed. It features pictures and audio in which it states the term as well as gives it context all in Dene. There is also an option to select audio for the English translation as well, however the explanation of the term is only aural (for both languages) and the context is not typed anywhere.

Another visual and audio based software was packaged by Harvey (2015), a Computer Science professor at the Southern University of Oregon. His *Acquisition of Restored Native Speech (ACORNS)* project software can be used by language teachers to build their own CALL tools. The software itself is language-independent and allows speakers to create a standard set of exercises for their own language. This program is comprised mostly of “hear and click” type questions which require the user to listen and select the picture that relates to the audio. Although most interaction involves pictures and audio, there is some text present. It would seem, however, that neither grammatical paradigms nor orthography are a strong focus for these exercises. Regardless, this easy to adapt language independent framework is a goal to work towards when developing CALL applications for endangered languages.

Currently, the source code for *nêhiyawêtan* is open source and free for language teachers to use. Although it can be made and adjusted to fit the needs of other languages, the software is not easily created for language communities without extensive knowledge of coding. Unfortunately the level of coding knowledge needed to create an application like *nêhiyawêtan* is much higher than the average language teacher possesses. The end goal would be to create an easy to use format for community language instructors to input course materials to create exercises without losing the focus on grammatical paradigms and relevant answer feedback.

Rosetta Stone has started to make programs for endangered languages as well. Iñupiaq, Innuttit, Mohawk, Navajo, and Chitimacha, all have worked with Rosetta Stone to create programs (Rosetta Stone 2015). Rosetta Stone is one of the mainstream CALL applications for language learning that promotes “technology-driven language, literacy and brain-fitness solutions (Rosetta Stone 2015). There is a demo available online for Navajo. Like the Ho Chunk application, the user is introduced to four pictures, however it differs in that each picture is labeled with the Navajo word above (Navajo Language Renaissance 2008). The audio describes each picture one at a time. Next a testing phase begins. The four pictures are then shown again and the audio plays with the target vocabulary listed in Navajo at the top of the page. The user subsequently selects the picture that it represents. After a lesson is finished, the user is given an evaluation of the number of correct and incorrect answers, as well as the number of those skipped and not seen. This is the first form of tracking feedback seen in any of these programs. The applications previously covered only features red “X”s and green check marks to indicate correct and incorrect answers, if at all.

Another minority language program is run by the First Nations Language Centre, headed by Marianne Ignace, Ph.D., at Simon Fraser University (SFU) and currently funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). SFU has been working

collaboratively with the Old Massett Village Council, and the Skidegate Haida Immersion Program (SHIP), in order to create a CALL program for Haida (Simon Fraser University 2015). What makes this project immediately different from the others is that the user must log in to the website. This allows the user, and their course instructor to track their progress. The login screen lets the user view which modules they have completed and their performance for each one.

The SFU program differs from the others again in that their vocabulary lists are completely in Haida and also have audio files for each term. This was not seen in any of the other applications or sites. The terms are explained in a tab located before the vocabulary list, where thorough definitions and their context of use are given. There is also the option to practice sentence construction. Each module has sentence structure thoroughly explained which the user can practice. Additionally, there is also a speaking practice task in which the user can record themselves and compare it to the proper pronunciation. Finally, there is translation practice. During this task, the user is meant to enter translations for the passages from Haida to English, or English to Haida. Although this program has a lot more focus on literacy than the other applications, programs and websites previously mentioned, it is still generally heavier on orality.

As described in the previous section, *nêhiyawêtan* is meant to free up class time to practice speaking the language by providing students the opportunity to complete word structure and paradigms drills on their own with feedback. It is also the first of the aforementioned publicly accessible applications to focus on literacy and grammar. This is a big difference from previous CALL endeavors for endangered languages that focus mainly on orality. By focusing on literacy, the program helps to create new domains for language use that orality does not.

## CHAPTER IV. METHOD

To reveal potential user problems, a self-study of *nêhiyawêtan* was conducted by the author (with English as a first language, learning Cree as a fourth language) before the user evaluations. The user evaluations will be the main focus of this thesis. The self-study was conducted to obtain insight on particular problems users might face when interacting with *nêhiyawêtan* in order to effectively prompt users to reveal their opinions and perceptions of these particular errors in the application. As previously mentioned, details and heuristic evaluation of the self-study are available in Appendix A.

### **4.1 Checklist**

During the design process a checklist motivated by Hagdahl (2000:33) was utilized. It highlighted “practical details” to consider when proposing a design. The checklist also included intentions for what would be done for each checklist item, and the motivation for each intention. This design provides a helpful format for the researcher to refer back to. Hagdahl’s (2000:33) checklist was inspired by Åkerstedt (1999), which would have been consulted as well, however, a translated version of the document was not available. It is hoped that this checklist will provide a starting point for future evaluations of CALL programs for endangered and indigenous languages. The complete checklist may be found in Appendix B.

### **4.2 Participants**

Five native English speakers, all of which are second language (L2) or third language (L3) learners of Plains Cree participated (4 female, 1 male). Four of the five students were L3 learners of Plains Cree. The fifth participant was an L2 learner. The other four learners had a varying L2 (e.g. French, Spanish, Latin). Three out of the five participants had visual impairments that were corrected by glasses. One out of the five participants had a learning disability regulated by medication and therapy. One participant was an L2 learner of Cree, while the other four were L3 learners. Regarding heritage language, three participants most strongly identified as Cree, while the other two most strongly identified as Canadian (though one of these participants was of Métis heritage).

The language acquisition ages, histories, and performance in Plains Cree are presented in Tables 1.1 and 1.2. The starting age of acquisition ranged from nineteen to twenty-seven years old. All five participants were sequential bilinguals. The education levels are fairly similar among these participants as all have completed some university; however, one of the participants was a graduate student (AVG=14.9 years of formal education, SD=2.7, range=13-19.5 years). Plains Cree exposure was reported by participants as being limited mostly to reading and language-labs/self-instruction. The participants communicated that the most important factors contributing to learning Plains Cree were interacting with friends and family, and reading.

Since *nêhiyawêtan* is based off of the NS 152 Introductory Cree course materials, only students who had taken NS 152 or higher were eligible to participate. Dorothy Thunder, the instructor of the Cree course, was contacted and she agreed to forward a recruitment email to her students. The students then made contact with the researcher directly and scheduled their participation for a convenient time. By showing interest through contacting the researcher, the students would not feel obligated to participate as they might have if required to respond to their instructor. Due to the limited participant pool, all participants were offered a small monetary compensation for their time.

Language History Measures	Cree L2 History <sup>a</sup>			Cree L3 History <sup>b</sup>		
	M <sup>c</sup>	SD	Range	M <sup>c</sup>	SD	Range
Self-reported proficiency <sup>d</sup>						
Understanding	4.0	N/A	N/A	3.8	1.7	2.0-6.0
Speaking	4.0	N/A	N/A	4.3	1.5	3.0-6.0
Reading	4.0	N/A	N/A	4.8	1.5	3.0-6.0
Age (yrs.)						
Started	20	N/A	N/A	22	3.8	19-27
Attained Fluency	N/A	N/A	N/A	N/A	N/A	N/A
Started Reading	20	N/A	N/A	22	3.6	19-27
Attained Fluency	N/A	N/A	N/A	N/A	N/A	N/A
Immersion Duration (yrs.)						
In a Country	20	N/A	N/A	23	3.5	20-27
In a Family	20	N/A	N/A	6.8	14	0.0-27
In School	0.2	N/A	N/A	1.0	1.4	0.1-3.0
Contribution to Language Learning <sup>e</sup>						
Family	4.0	N/A	N/A	4.5	4.2	0.0-10
Friends	5.0	N/A	N/A	6.3	4.4	0.0-10
Reading	5.0	N/A	N/A	7.3	2.1	5.0-10
T.V.	0.0	N/A	N/A	0.5	1.0	0.0-2.0
Radio	0.0	N/A	N/A	0.3	0.5	0.0-1.0
Independent Study	9.0	N/A	N/A	3.8	3.4	1.0-8.0
Extent of Language Exposure <sup>f</sup>						
Family	2.0	N/A	N/A	1.8	1.5	0.0-3.0
Friends	2.0	N/A	N/A	0.8	1.0	0.0-2.0
Reading	5.0	N/A	N/A	3.5	1.0	3.0-5.0
T.V.	0.0	N/A	N/A	0.8	1.0	0.0-2.0
Radio	0.0	N/A	N/A	0.5	0.6	0.0-1.0
Independent Study	7.0	N/A	N/A	3.8	4.5	0.0-10
Self-Report of Foreign Accent <sup>g</sup>						
Perceived by self	4.0	N/A	N/A	6.5	3.9	1.0-10
Identified by others	1.0	N/A	N/A	9.3	1.5	7.0-10

<sup>a</sup>Cree L2 History: 1 Participant. <sup>b</sup>Cree L3 History: 4 Participants. <sup>c</sup>M: mean. Ranges N/A for L2 since it is only the data from one person. <sup>d</sup>range: 0 (none) to 10 (perfect). <sup>e</sup>range: 0 (not a contributor) to 10 (most important contributor). <sup>f</sup>Range: 0 (never) to 10 (always). <sup>g</sup>Range: 0 (none) to 10 (pervasive).

Table 1.1: Self-reported language history and proficiency data for Plains Cree extracted from LEAP-Q.

Language History Measures	NS 152 <sup>a</sup>			NS 252 <sup>b</sup>		
	M <sup>c</sup>	SD	Range	M <sup>c</sup>	SD	Range
Self-reported proficiency <sup>d</sup>						
Understanding	4.7	1.2	4.0-6.0	2.5	2.1	2.0-3.0
Speaking	4.0	1.0	3.0-5.0	4.5	1.5	3.0-6.0
Reading	4.3	1.5	3.0-6.0	5.0	1.4	4.0-6.0
Age (years)						
Started	22	4.4	19-27	21	2.1	19-22
Attained Fluency	N/A	N/A	N/A	N/A	N/A	N/A
Started Reading	22	4.0	20-27	21	2.1	19-22
Attained Fluency	N/A	N/A	N/A	N/A	N/A	N/A
Immersion Duration (years)						
In Country	22	4.0	20-27	22	3.1	20-24
In a Family	16	14	0.0-27	0.0	0.0	N/A
In School	0.1	0.1	0.1-0.2	1.9	1.5	0.8-3.0
Contribution to Language Learning <sup>e</sup>						
Family	5.7	3.8	3.0-10	2.5	3.5	0.0-5.0
Friends	7.7	2.5	5.0-10	3.5	5.0	0.0-7.0
Reading	7.3	2.5	5.0-10	6.0	1.4	5.0-7.0
T.V.	0.7	1.2	0.0-2.0	0.0	0.00	N/A
Radio	0.3	0.6	0.0-1.0	0.0	0.00	N/A
Independent study	7.3	2.1	5.0-9.0	1.0	0.00	N/A
Extent of Language Exposure <sup>f</sup>						
Family	2.0	1.0	1.0-3.0	1.5	2.1	0.0-3.0
Friends	1.3	1.2	0.0-2.0	0.5	0.7	0.0-1.0
Reading	3.7	1.2	3.0-5.0	4.0	1.4	3.0-5.0
T.V.	0.7	1.2	0.0-2.0	0.5	0.7	0.0-1.0
Radio	0.3	0.6	0.0-1.0	0.5	0.7	0.0-1.0
Independent Study	4.0	3.0	1.0-7.0	5.0	7.1	0.0-10
Self-Report of Foreign Accent <sup>g</sup>						
Perceived by Self	4.0	3.0	1.0-7.0	9.0	1.4	8.0-10
Identified by Others	6.0	4.6	1.0-10	10	0.0	N/A

<sup>a</sup>NS 152: 1 Participant. <sup>b</sup>NS 252: 4 Participants. <sup>c</sup>M: mean. Ranges N/A for L2 since it is only the data from one person. <sup>d</sup>range: 0 (none) to 10 (perfect). <sup>e</sup>range: 0 (not a contributor) to 10 (most important contributor). <sup>f</sup>Range: 0 (never) to 10 (always). <sup>g</sup>Range: 0 (none) to 10 (pervasive).

Table 1.2: Self-reported language history and proficiency for participants in Study. Two participants were in NS 252, while the other three were in NS 152. The data shown was extracted from LEAP-Q.

### 4.3 Study Outline

In addition to the checklist, Hagdahl (2000:34) created a study outline, also based on a guide in Åkerstedt (1999). The original guide had been created by Åkerstedt (1999) for an interview, but has since been updated by Hagdahl (2000:34) to include user tasks and an observation period. This guideline has been further edited from Hagdahl (2000:34) to include the issuing of the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al. 2007), and a user evaluation questionnaire designed specifically to gain insight on user perceptions of *nêhiyawêtan*.

The user evaluation questionnaire has replaced Hagdahl's interview. Since many of these questions will be touched on during the practice portion of the session, this will give the participant more freedom to discuss the weaknesses of the application without having to confront the researcher outright. Overall, this guideline was a useful tool for ensuring all the necessary information was given and collected throughout the study. It is hoped that this will be a helpful guideline for others evaluating CALL programs for endangered and indigenous languages.

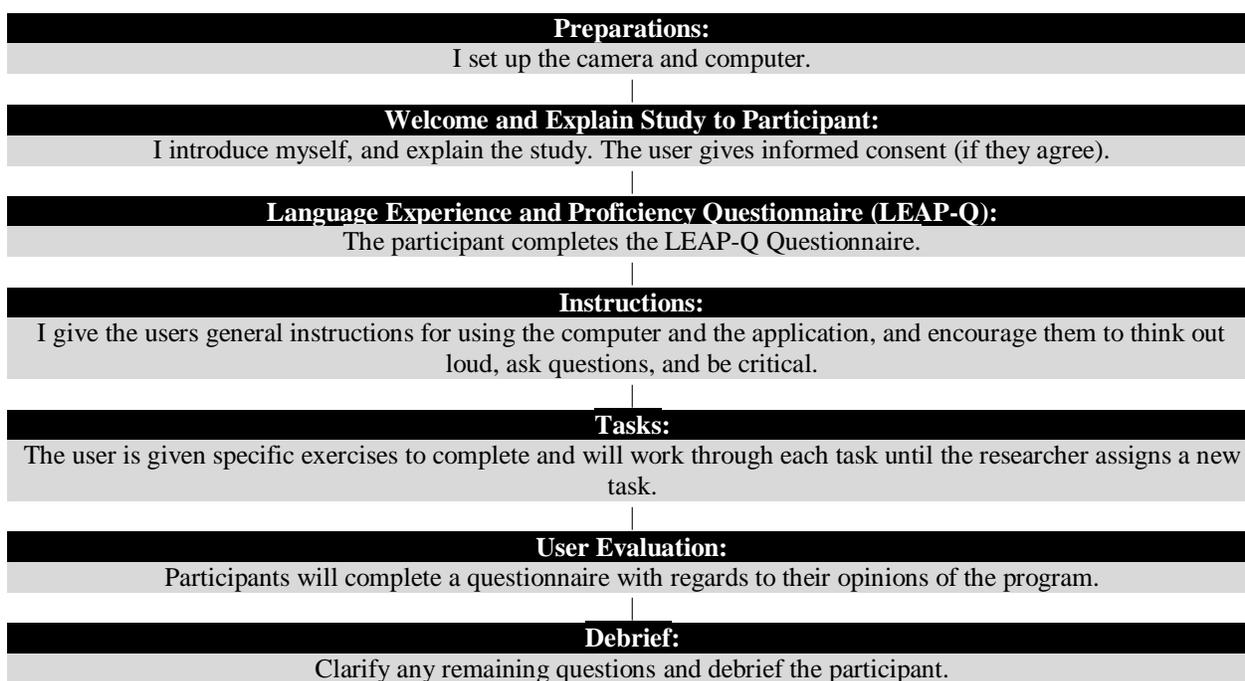


Figure 2.1 : The study outline, motivated by Hagdahl's (2000:34) study guideline.

Below is a brief explanation of each item in the outline in Figure 2.1.

#### 4.31 Preparations

Monk et al. (1993:7) had a useful checklist for setting up cooperative evaluation studies which was helpful in preparing the session environment. Set up involved running QuickTime Player Version 10.4 (833.7) on a MacBook Pro (13-inch, Mid 2012). For the screen-casting, QuickTime Player was set to use the internal microphone, and to show mouse clicks. A video camera was also used to record three out of the five sessions. The set up took approximately fifteen minutes.

The video camera was used initially as a precautionary measure (in case the screen recording failed), and was only focused on the screen. After the first few sessions, it was clear that the additional camera was unnecessary. Additionally, the files were bulky and inefficient. Eliminating the video camera simplified the study set-up. Without the video camera and tripod, set-up took around five minutes. The last two sessions relied only on the QuickTime screen-casting, and the absence of the video camera did not seem to have any adverse effects on the study. The three participants that were recorded by the video camera all indicated on the user evaluation questionnaire that they did not find the recording equipment intrusive; therefore, this should not affect the results.

Four out of five sessions took place in a large custom built soundproof booth in the Alberta Phonetics Laboratory (APhL) in the Department of Linguistics at the University of Alberta. For the final session the booth was unavailable. Instead, the final session was conducted in a meeting room located between faculty offices, which made for a silent environment with no distractions. The main difference between the two spaces is the size of the room, which should not have a large effect on the data collected.

One small change was made to the initial set-up of the study after the first participant requested scrap paper. This practise then became part of the set-up. Giving the students scrap paper made the environment similar to what they might experience outside of class time. Out of the five users, three opted to use the scrap paper.

#### ***4.32 Welcome and Explain Study to the Participant***

Due to the previous email correspondence with each participant, introductions were relatively brief. The researcher explained to each participant that they would be working through a series of exercises and answering questions pertaining to grammatical paradigms using an online CALL application. The origin and intention of the application was explained briefly. Participants were also informed of the various recording equipment present in the laboratory and were ensured that it would consist of audio and screen casting only. The minimal risks were explained and the participants were able to give informed consent. The participants were also encouraged to think out loud and instructed to express any thoughts they had on the application or any problems they encountered.

#### ***4.33 Language Experience and Proficiency Questionnaire (LEAP-Q)***

The LEAP-Q was chosen because it attempts to represent multiple factors of the status of multilingualism. For example, it incorporates questions of language competence, age of acquisition, acquisition environment, language exposure, and language use (Marian et al. 2007). The participants were given the Canadian version which was adapted to pencil-and-paper by Marilyn Logan. The study favoured the pencil-and-paper version over the digital copy since the participants would already be required to spend an extended period of time working on the computer. It was hoped that this would allow the participants to better focus on the tasks for which computer use was essential. This questionnaire added quantitative data to the study.

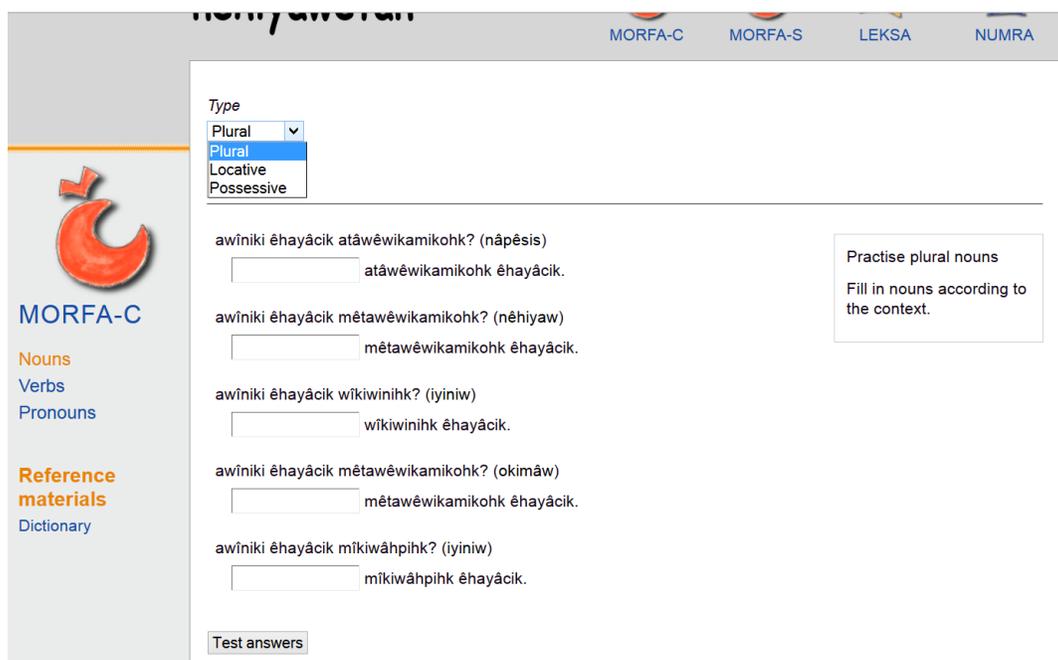
#### ***4.34 Instructions***

Participants were again reminded to verbally express their thoughts as naturally as possible and given brief instructions to make the tasks clear. The participant was presented with the *nêhiyawêtan* start page, as shown in Figure 1.1. From here, they were given the exercise type (e.g.

Morfa-S), the part of speech (POS), and the type of inflection that would be practiced (e.g. plural, past, etc.). The participant was given this simple instruction to gauge the navigability of the user interface. Once the user reached the correct exercise, they were instructed to follow the instructions given by the application.

### 4.35 Tasks

Each participant went through the same set of exercise types. Morfa-S (plural and diminutive nouns), Morfa-C (plural, locative, and possessive nouns), and Leksa (Cree to English and English to Cree words) were practiced for five minutes per inflection or translation type. Morfa-S (past tense verbs) and Morfa-C (present tense verbs) were assigned for 10 minutes each. Greater time was allotted for verb inflection tasks since they generally were more difficult for users and took longer for users to complete a set. The tasks were stated aloud after each exercise was complete, so that the user would not be worried about the amount of exercises they had to complete. Within each exercise the user was free to use the application as they pleased. This would provide the researcher with insight into user strategy.



The screenshot shows the Morfa-C interface. On the left, a navigation menu includes 'MORFA-C' (selected), 'Nouns', 'Verbs', 'Pronouns', 'Reference materials', and 'Dictionary'. The main content area features a 'Type' dropdown menu with 'Plural' selected. Below this, there are five practice questions, each with an input field and a label: 'awīniki éhayá cik atáwéwīkamīkohk? (nápésis)', 'awīniki éhayá cik métawéwīkamīkohk? (néhiyaw)', 'awīniki éhayá cik wīkiwīnihk? (iyiniw)', 'awīniki éhayá cik métawéwīkamīkohk? (okimáw)', and 'awīniki éhayá cik mīkiwáhpīhk? (iyiniw)'. A 'Test answers' button is at the bottom left. A callout box on the right contains the text: 'Practise plural nouns' and 'Fill in nouns according to the context.'

Figure 2.2: Morfa-C plural noun inflection practice. The user selects the inflection type from the drop-down menu, while the POS is selected from the side menu.

Although there were times designated for each exercise type, the actual time spent by participants on each type varied. As a result of a limited wordlist (264 nouns, 182 verbs), some of the sessions began repeating words. It bored the participants to continue repeating words they had previously completed. For this reason the time spent on some exercises during some sessions was reduced. Exercise timeframes were also affected by the nature of the exercises. Some users may have finished a set of questions early, but not early enough to complete another set. Others may have gone over time wanting to finish self-correcting their answers before moving on.

Despite the differences in time spent, most participants completed the same number of each type of exercise. These tasks were chosen because it provides the user with a full complement of all the available functions that users may opt to use. Each of these exercises has different help feedback given by the “help” function. Testing each exercise will provide video documentation on how the users understand this feedback.

#### ***4.36 User Evaluation***

The user evaluation questionnaire was issued immediately after the CALL session. The questionnaire went over functions that the user had seen during the session, and asked for their opinion. Participants were given the opportunity to express their perceptions of the pros and cons of the application, as well as what they would like to see for the application in the future. This was a useful way of discovering what prospective users found the most rewarding or troubling within the application. Most of the questions were specific to *nêhiyawêtan*, although some of the more general questions were inspired by Monk et al. (1993:11). This questionnaire can be found in Appendix C.

#### ***4.37 Debrief***

When debriefing, participants were reminded of the purpose of the study. Afterward, any of the researcher’s remaining questions were clarified by the participant, and vice versa. The participants were then given their compensation and the study was concluded.

## CHAPTER V. RESULTS

After each session, the videos resulting from the screen cast were viewed and user actions were meticulously recorded. The user perceptions from the survey that concluded each session were also documented. The number of participants with shared experiences and opinions were documented and are displayed in Tables 2.1-2.8.

### 5.1 Navigation

Phenomena observed	# of users
1. Users did not pay attention to the interface language option on the homepage.	5 users
2. User was able to navigate the main home page options with ease.	5 users
3. User had trouble finding the instructions for the exercises.	4 users
4. User was able to find the instructions almost immediately.	1 user
5. User suggested moving the instructions to a more noticeable location.	2 users
6. User had trouble finding where to change the POS	2 users
7. User was able to find where to change the POS almost immediately.	3 users
8. User did not notice the dictionary link.	5 users
9. User was able to find where to change the type of exercise almost immediately.	5 users
10. User understood how to navigate from one exercise to another without using the home page.	5 users
11. User requested better presentation and navigation of the interface.	2 users
12. User found the 'click on a word' function by accident.	5 users

Table 2.1: *The number of participants that experienced or perceived observed phenomena while navigating nêhiyawêtan.*

The results in Table 2.1 pertain to the navigability of *nêhiyawêtan*. These phenomena were observed to gauge how easily users can find various functions of the application, as well as various links and general website navigation. More general observations of the design, structure, and content of the interface and exercise templates overall were also documented. These phenomena are show in Table 2.2.

## 5.2 General Interface Design & Template Structure

Phenomena observed	# of users
1. User noticed that some of the vocabulary, content, and derived inflectional forms were not consistent with the textbook.	4 users
2. User requested having the option to review the paradigm before being tested on it.	1 user
3. User suggested colour coding, or different fonts to make questions and instructions more clear.	2 users
4. User enjoyed the all the Cree sentence contexts.	1 user
5. User requested translation and analyzation exercises.	1 user
6. User requested more audio.	1 user
7. User did not use audio feature when available.	1 user
8. User used the audio feature when available.	2 users
9. User never encountered the audio feature.	2 users
10. User explicitly stated that they enjoyed the immediate feedback, and ability to explore the language by self-correcting outside of class time.	2 users
11. User explicitly stated that they enjoyed having multiple tries to get the correct answer.	1 user
12. User thought the application adequately represented the course material.	1 user
13. User thought the application should better follow the course material.	2 users
14. User was frustrated by repeating questions.	1 user
15. User enjoyed having a variety of words to practice with.	2 users
16. User requested sample answers for each exercise.	1 user
17. User did not like encountering unfamiliar vocabulary not yet covered in class.	2 users

Table 2.2: *The number of participants that experienced or perceived observed phenomena of nêhiyawêtân generally.*

## 5.3 Observing Use of Functions

In addition to more general observations and evaluations, phenomena specific to the various functions of *nêhiyawêtân* were also observed. The phenomena observed for the ‘click on a word,’ ‘help,’ ‘new set,’ and ‘test answers’ functions are shown in Tables 2.3, 2.4, 2.5, and 2.6 respectively. These tables outline the usage and user opinions of each of these functions.

Phenomena observed	# of users
1. User made use of the 'click on a word' function.	5 users
2. User found the 'click on a word' function useful.	5 users
3. User requested 'click on a word' function to be available for inflected forms and sentences as well as base-forms in questions.	2 users

Table 2.3: *The number of participants that experienced or perceived observed phenomena of the 'click on a word' function.*

Phenomena observed	# of users
1. User made use of the 'help' function.	5 users
2. User found the 'help' function helpful.	3 users
3. User found the 'help' function initially unclear.	1 user
4. User requested shorter 'help' explanations.	1 user

Table 2.4: *The number of participants that experienced or perceived observed phenomena of the 'help' function.*

Phenomena observed	# of users
1. User used the 'new set' function because s/he thought the given question set was too difficult.	3 users
2. User used the 'new set' function because the questions were repeating themselves.	1 user
3. User only used the 'new set' function after completing a set of questions.	2 users

Table 2.5: *The number of participants that experienced or perceived observed phenomena of the 'new set' function.*

Phenomena observed	# of users
1. User used the 'test answers' function more than once per set.	5 users
2. User used the 'test answers' function once, and then immediately used the 'show answers' function.	5 users
3. User used the 'test answers' function and all of their answers were correct. They did not bother using 'show answers.'	4 users
4. User used the 'test answers' function and all of their answers were marked as 'correct' though they contained errors. They did not bother using 'show answers' and in turn were not shown their errors.	3 users
5. User used the "test answers" function and all of their answers were marked as correct. They used "show answers" anyway.	2 users
6. User requested the exact location of the error be highlighted in red when the 'test answers' function is used to show the user what they need to improve.	1 user

Table 2.6: *The number of participants that experienced or perceived observed phenomena of the 'test answers' function.*

### 5.4 Evaluating Verb Exercise Templates

Along with the various functions, the phenomena that were observed and experienced by users while working on the verb exercises were also documented. This provides insight into the specific difficulties users faced while working through these particular exercises, whether due to content or template structure. Additionally, user suggestions for improvement were included.

Phenomena observed	# of users
1. User did not read the context of the questions in Morfa-C.	4 users
2. User found the formatting of the Morfa-S verb exercises confusing.	5 users
3. User requested to be given the option to work from English to Cree in the verb exercises.	1 users
4. User did not understand the vocabulary in the sentences in Morfa-C.	1 user
5. User requested an exercise for practicing conjunct verbs.	1 users
6. User requested the option to work on transitive inanimate and animate intransitive verbs in separate exercises.	1 user

Table 2.7: *The number of participants that experienced or perceived observed phenomena during verb exercises.*

### 5.5 Error Analysis

During this observational period, the participant errors were also tracked. This will provide insight as to how well the user errors match the help given by the ‘help’ function. The number of participants that produced each error are recorded in Table 2.8.

Errors	# of users
1. Error related to vowel length.	4 users
2. Error was a close misspelling of the word.	5 users
3. Answer was left blank.	5 users
4. User produced the wrong inflected form.	5 users
5. User produced the right suffix/prefix but implemented it in the wrong way.	5 users
6. Error due to the change from t-c in the diminutive form.	2 users
7. Error reversing the diminutive form when the baseform of the noun ends in a vowel.	1 user
8. Error due to inflexibility of accepted answers in Leksa.	3 users
9. Error due to user not knowing vocabulary.	3 users
10. Incorrect answer, but related semantically.	4 users
11. Error due to missing t-joiner.	1 user
12. Error due to missing h-joiner.	2 users

Table 2.8: *The number of participants that produced each error while using nêhiyawêtân.*

The frequency of errors was also recorded (Figure 3.1). It is hoped that this data will assist in prioritization of help messages. By analysing these errors, the anticipated help messages can be made more relevant.

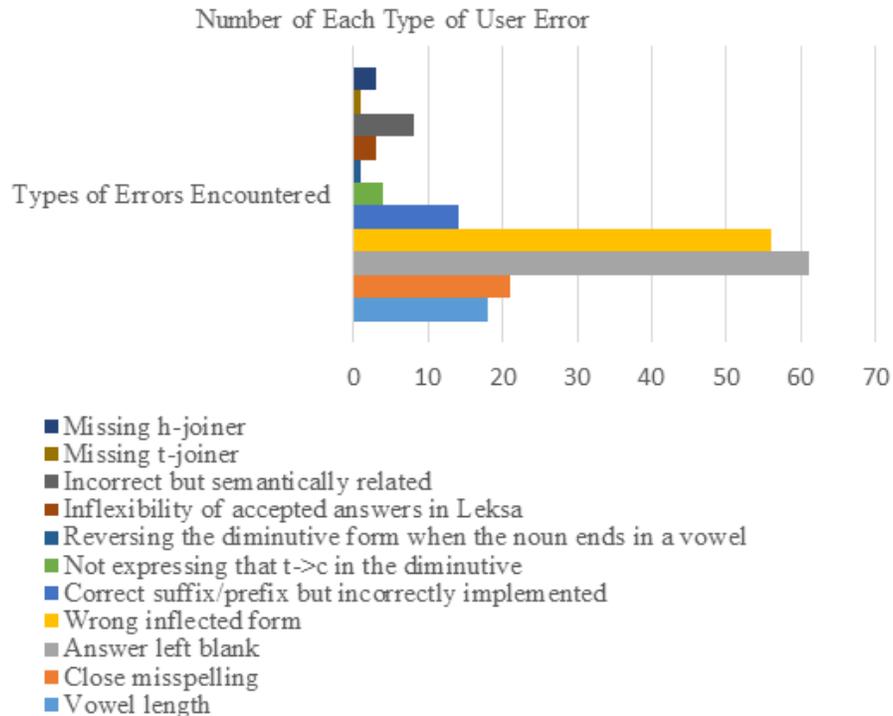


Figure 3.1: *Frequency of error types.*

## 5.6 User performance

In addition to tracking how many users experienced, perceived or produced certain phenomena, the number of times each phenomenon occurred was also recorded. Firstly, Table 3.1 describes the approximate scores achieved by users for each question set in each exercise. Although individual performance was not a focus of this study, this table can provide an intuition as to which exercises were the most difficult for participants.

Question Type	Questions Set Scores (Of 106 Sets)		
	<sup>a</sup> M	SD	<sup>b</sup> Range
Morfa-S Plural	2.45	1.75	0.00-5.00
Morfa-S Diminutive	2.67	1.94	0.00-5.00
Morfa-C Plural	4.78	0.44	4.00-5.00
Morfa-C Locative	4.50	0.71	3.00-5.00
Morfa-C Present	3.83	1.40	0.00-5.00
Morfa-S Past	1.00	0.82	0.00-2.00
Morfa-C Possessive	4.42	0.79	3.00-5.00
Leksa Cree-to-English	2.87	1.13	1.00-4.00
Leksa English-to-Cree	1.86	1.03	1.00-4.00

<sup>a</sup>M: mean. <sup>b</sup>Range: 0 to 5 (full marks).

Table 3.1: *Participant scores for each exercise type.*

### 5.7 ‘Show Answers’

The amount of times users did and did not use the ‘show answers’ function was also collected. *nêhiyawêtan* on occasion allows for a single vowel length error within a word, multiple errors are only allowed when the errors are all made regarding the same character (i.e. instead of producing “*kwâpahôpâkanak*” the user produces “*kwapahôpakanak*” - which has two errors both regarding “*a*” being produced instead of “*â*”). The application does this by marking the answers in green, meaning that they are correct, in all other respects except vowel length marking.

The amount of times participants used the ‘show answer’ function in various scenarios was recorded. In Table 3.2, “all answers were correct” is referring to the participant getting the answers all marked correct, whether it be on the first try or after self-correcting. “Some self-correcting” is when the participant does not get all of the answers marked ‘correct’ but has still made an attempt to correct the answers. “Right away” refers to those participants that peek at the answer immediately after typing in their first answer. This final action involves no self-correcting.

	% of all sets (of 106 sets)
User showed answers	
When all answers were ‘correct’	7%
After some self-correcting	18%
Right away	31%
User did not show answers	
All answers ‘were correct’	40%
Vowel length error	3%
Word never corrected/entered	1%

Table 3.2: *Percentages of the number of question sets during which the participant used or ignored the ‘show answers’ button.*

### 5.8 ‘Help’ Feedback

On average, participants completed 22 question sets each. The frequency of use of the ‘help’ function was also documented (see Table 3.3). Additionally, the frequency of on-topic help and off-topic help was looked at to ensure that users were provided with adequate feedback that allowed them to self-correct their answers. The ‘help’ that “would have been helpful” refers to help messages in which content addressed each user’s error. However, it is also possible that the user could have read the available help and still have been confused. Because we have no way of knowing this, Table 3.3 examines on-topic versus off-topic help.

The ‘help’ function being “Some help” refers to the occurrence of multiple errors in a users’ production. This is where one error might be addressed by the ‘help’ function, while the other is not; for example, when inflecting nouns in the plural form, a user enters “*soniyawikamika*”, when the correct answer is “*sôniyâwikamikwa*.” This user’s production contained both vowel length errors and an inflection error, when only the inflectional error is addressed by the given ‘help’.

% of all times help was available (101 times total)	
<b>Used help</b>	
Helpful	28%
Some help	4%
Not helpful (on topic)	15%
Not helpful (does not address the problem)	10%
<b>Did not use help</b>	
Would have been helpful	35%
Would have been some help	0%
Would not have been helpful	8%

Table 3.3: *Percentages of the number of times the ‘help’ function was available, whether or not users made use of it and whether or not it was helpful.*

### 5.9 Self-Correction

The amount of times users took initiative to self-correct their answers was also documented (see Table 3.4). When “user completely corrected answers” the user self-corrected to the point where all answers were all highlighted in green to indicate that they were correct. The user was considered to have “attempted to correct answers” if they had at least read the ‘help’ function or re-submitted a couple answers without being able to self-correct the answers completely. The times the “user did not attempt to correct answers” were when users immediately used the ‘show answers’ function.

Extent of Self-Correcting	% of initial errors (96 errors)
User completely corrected answers	52%
User attempted to correct answers	27%
User did not attempt to correct answers	18%
User left the question blank.	3%

Table 3.4: *Percentage of the number of times users self-corrected and to what extent. This table excludes errors made in Leksa exercises since there is no ‘help’ function available.*

## CHAPTER VI. DISCUSSION

### ***6.1 Navigation***

The occurrences of the phenomena in Table 2.1 showed that all five participants did not pay attention to the drop-down list on the homepage which allows users to change the interface language. However, they did all manage to easily navigate the other homepage options. This is most likely attributed to the irrelevance of changing the interface language for the tasks that were assigned to the participants. None of the participants had trouble switching from one exercise to another. They were able to locate each type of exercise relatively quickly. Navigating within exercise types, however, was seemingly more difficult. Two out of the five users had trouble finding where to change the part of speech which is located on the side menu. This could potentially be confusing because the majority of drop-down choices for the exercises are listed above the exercises. The part of speech is the only option listed on the side menu.

### ***6.2 ‘Click on a word’***

Since two users did not like encountering vocabulary that was new to them and two enjoyed learning the new vocabulary, the ‘click on a word’ function was very well received by the participants. All five participants made use of it and found it helpful. One participant was unable to understand the vocabulary used in the sentences in the Morfa-C questions and could have benefited having their definitions. Although, two participants requested that the ‘click on a word’ function be available for more words (both inflected and non-inflected), none of the participants used the dictionary link on the side menu. This is most likely because the participants did not notice the link while working through exercises. It could also be that it is burdening for a user to have to leave the interface for help when learning a new language, especially since the ‘click on a word’ function demonstrates that the same thing can be effectively achieved within the application. This function could be easily improved by using an existing plug-in linked to the dictionary or a similar functionality which does not require leaving the interface.

Moreover, all five users found the function by accident, which suggests deficient instruction. The participants either accidentally clicked on a word, or noticed that they could click on the word as the cursor changed as they scrolled over it. Although it is encouraging that this function is relatively easy to find without instruction, it is concerning that the instructions are not effective.

### ***6.3 Instructions***

Four out of the five users had difficulty finding the instructions. This is most likely because they are located to the right of the questions instead of above or to the left. This is inconvenient for native-English speakers because they are accustomed to reading from top to bottom and from left to right. The order of these features in the interface does not directionally adhere to this convention. Two of the five participants suggested moving the instructions to a more noticeable location. It was also suggested by two participants that the instructions should make use of bolded or italicized words and colour coding to draw the users attention. Another user suggested including a sample answer within the instructions to give the user a better idea of what is being asked of them.

#### **6.4 The ‘help’ Function**

The ‘help’ function was easily located and used more commonly across users. All five users used the ‘help’ function multiple times during each session. Two users explicitly stated that they enjoyed the option for immediate feedback outside of class time. However, three users found it helpful, while the other two users thought that the terminology and length made the ‘help’ confusing at times. One of the five users explicitly asked for shorter ‘help’ explanations. This could be possible if feedback could be given gradually. Heift (2001) explained that prioritization is necessary to give users feedback when multiple errors are present. The downside to this is that users only correct the errors that are explained by the system and ignore all other existing errors (Heift 2001). On the other hand, one of the five participants that found the help confusing had requested that the given help better lead the user to the correct answer. Unfortunately, more obvious user feedback may not be the best way to generate long-term retention of inflectional knowledge.

Bell et al. (2008) discusses “desired difficulties.” These difficulties occur when learners are required to do work that is more immediately mentally taxing but has long-term retention benefits. It is important to note that when looking at desired difficulties for *nêhiyawêtan*, we are more focused on difficulties due to content, not question format. It is less critical for learners to retain knowledge of how to complete a particular type of question. Moreover, the desired difficulties should be content-related in order to enhance long-term retention of inflectional knowledge of Plains Cree in users. Therefore, if the participant understood the instructions but had to work harder to self-correct their answer, it might not necessarily be a feature that needs to be changed extensively. Especially since the most commonly observed errors (excluding Leksa results where help is not available) were inflectional, which are addressed in the current ‘help’ messages. However, one user did suggest to have a review of the paradigms before attempting the questions, which could potentially be helpful. This could be resolved by creating a parallel online coursebook for reference, although current students could simply resort to their coursebooks until such a resource is implemented.

#### **6.5 ‘Test answers’ and ‘Show answers’**

All five participants made use of the ‘test answers’ function more than once per set of questions; answers must be tested at least once to be able to ‘show answers’. One participant even noted that they really enjoyed being able to submit answers multiple times. All five participants, some more often than others, would also occasionally ‘show answers’ immediately after their first attempt. This phenomena was most common for Leksa question sets. When being tested on vocabulary, it is harder to guess the answer and there is no help function in place for this exercise. This most likely attributed to the more frequent early use of the ‘show answers’ function.

Currently, *nêhiyawêtan* allows for some deviations in vowel length. Unfortunately, the user is not made aware of these errors effectively (see Figures 4.1-4.2).

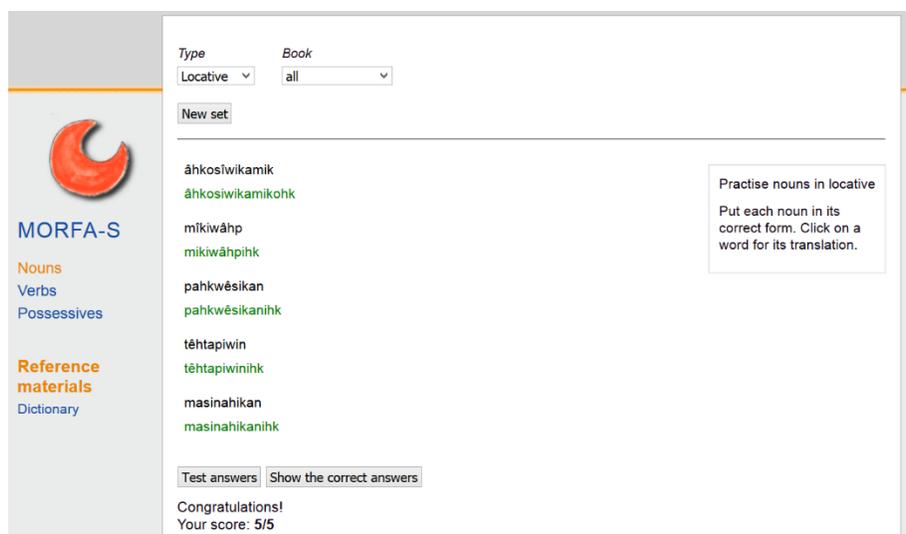


Figure 4.1: *Morfa-S Locatives*. This is what users see when they submit their answers with single vowel length errors. The first and second entries should be “âhkosiwikamikohk” and “mikiwâhpihk” respectively. Both answers are missing the length marker over the ‘i’, but at a glance the answers appear correct to the user.

In order to see that an error has been made concerning vowel length, the user needs to ‘show answers’ after submitting their answers even if they have all been marked as correct (see Figure 4.2). Out of the five participants only two did this at least once during their session. Even still, the error is not made obvious unless closely comparing produced answers to expected answers. Ideally vowel length errors would be marked in another colour, such as yellow, when the user ‘tests’ answers. Moreover, users should be shown their errors and given the opportunity to correct them.

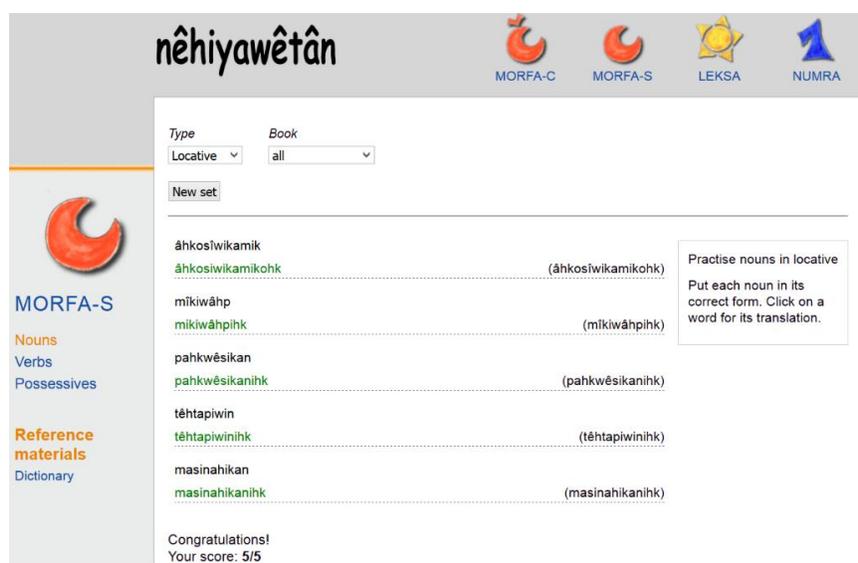


Figure 4.2: *Morfa-S Locatives*. After the user selects ‘show answers’ they are able to see that their answers differ slightly from the correct answers; however, they are not overtly notified.

Although, four out of the five users did not ‘show answers’ at least once when all of their answers were ‘correct’, only three users missed errors because of it. To counter this, one participant suggested that the exact location of the error be highlighted in all erroneous productions. It would also be ideal if this were to happen after answers are tested and users were given the opportunity to correct their errors. Currently, users are unable to correct these vowel length errors once submitted (see Figure 4.2).

### **6.6 ‘New Set’**

After completing each set of questions the user must select ‘new set’ to access new questions. Two users used this function for this reason only. The other three users also used it to explore new options. They used ‘new set’ to avoid question sets that were too difficult. This mostly pertains to Leksa, as it is harder for the user to improvise answers when they have not learned the vocabulary. This is consistent with the results of Heift & Rimrott (2012) in which avoidance was triggered mostly by meaning-based questions. In Heift & Rimrott (2012) users were able to avoid using words by using synonyms or separate phrases in essay questions. However, because of the rigidity of the Leksa vocabulary exercises, lexical choice was limited and avoidance using synonyms were not an option. In order to avoid getting these questions incorrect, participants were forced to use the ‘new set’ function. One of the three users also used it to avoid repeating questions, which caused frustration for this learner. Due to the relatively small vocabulary covered in the textbook material, further restrained by the context of questions, there are often words that are repeated. However, only two participants really took note of this. Another possibility for future versions of *nêhiyawêôtân* is a function for tracking words seen by users and only producing new words; however, this would require a login function as well.

### **6.7 Audio**

The application contains a small amount of audio files linked to vocabulary. These files are only available in Leksa for certain words, however future versions of *nêhiyawêôtân* intend to include more options for audio. Only three of the five participants had the opportunity to use the audio files. Out of these three only one participant did not use the audio feature. One of the users that did not have the opportunity to use the function requested more audio. Adding more audio will not only increase knowledge of proper pronunciation for learners, but also provide fluent speakers with a reference to familiarize themselves with the orthography and increase literacy.

### **6.8 Content Relevance**

The main goal of this application was to provide students of the NS 152 Introductory Cree course with immediate answer feedback on class material outside of class time. Three of the five students thought that the application adequately represented the course material, while the other two thought it could better follow the course material. This discrepancy is due to the inconsistency of vocabulary, content and derived inflectional forms in relation to the textbook. Although four users observed these differences, two were strongly affected by the variance. It is important for learners to have consistency when learning a new language, and users can find inconsistencies frustrating. Updated versions of *nêhiyawêôtân* will attempt to eliminate these variances to streamline learning for new users.

## 6.9 Question Formatting and Interface Options

Due to the formatting of some questions, the users' process of answering these questions was different than intended. For example, for all of the Morfa-C exercises four out of five participants did not read the context of the questions. The idea behind Morfa-C was to give users the opportunity to practice inflection within the context of a sentence. Unfortunately, none of these sentences actually requires the user to understand any of the vocabulary (see Figure 4.3). These users only paid attention to the part of speech that was required to answer the question.

The screenshot shows the Morfa-C interface. On the left is a sidebar with the Morfa-C logo (a red crescent moon) and navigation links: Nouns, Verbs, Pronouns, Reference materials, and Dictionary. The main area has a 'Type' dropdown set to 'Plural' and a 'New set' button. Below are five questions, each with a text input field:

- awíniki é-ayáćik ocenásisihk? (ayisiyiniw)  
ocenásisihk é-ayáćik
- awíniki é-ayáćik ocenásisihk? (nápésis)  
ocenásisihk é-ayáćik
- awíniki é-ayáćik atáwékamikohk? (nápéw)  
atáwékamikohk é-ayáćik
- awíniki é-ayáćik áhkosiwikamikohk? (iskwésis)  
áhkosiwikamikohk é-ayáćik
- awíniki é-ayáćik atáwékamikohk? (awásis)  
atáwékamikohk é-ayáćik

At the bottom is a 'Test answers' button. On the right, a text box says: 'Practise plural nouns. Complete the response to the question by entering the plural form of the noun given in brackets.'

Figure 4.3: Morfa-C. An example of the context based questions. The user can inflect the noun (in parentheses) for plural without reading the context.

The exercises that the participants were most displeased with were the Morfa-S verb exercises (see Figure 4.4). All five participants found the exercise template to be confusing.

The screenshot shows the Morfa-S interface. On the left is a sidebar with the Morfa-S logo (a red crescent moon) and navigation links: Nouns, Verbs, Possessives, Reference materials, and Dictionary. The main area has 'Tense' set to 'present' and 'Animacy' set to 'AI-TI'. Below is a 'New set' button and five questions, each with a text input field:

- nácam  
(wíya) anohc  nípiy
- itohtatáw  
(kíya) anohc  tohtósápoy
- kinaskwénikonaw  
(níyanán) anohc
- otápásow  
(kíya) anohc
- kinítawéyíimikonaw  
(kíyánaw) anohc

At the bottom is a 'Test answers' button. On the right, a text box says: 'Practise present tense. Enter the above verb in the form that corresponds to the pronoun in brackets. Click on a word for its translation. Note: All verbs are given in the 3S form.'

Figure 4.4: Morfa-S present tense verb exercise.

In Figure 4.4, each word above the line with the answer slot is the verb the user is intended to inflect. The verb is inflected for person dependent on which pronoun is featured in parentheses. Transitive inanimate (TI) verbs have direct objects after the answer slot to distinguish them from animate intransitive (AI) verbs. Currently, there is no option to practice AI and TI verbs separately. One user had requested this. This option will undoubtedly be developed for future versions of *nêhiyawêtan*.

One user also requested exercises for practicing other verb paradigms, especially conjunct paradigms. This was intended, but for time's sake was not implemented in the demonstration version. Adding conjunct and other excluded verb paradigm exercises (such as imperative forms) would also contribute to a better complement of course material.

Other suggested exercises include translation and analyzation exercises. This is relevant to the course curriculum and much class time is devoted to practicing these. Translating and analysing sentences is also thoroughly tested within class time. Having these types of questions would help *nêhiyawêtan* to better accompany the course material.

With regards to interface options, two users pointed out that the interface is not especially culturally relevant. They also would prefer an updated interface that is more appealing. Although this is important, there are other short-comings that must first be addressed.

## CHAPTER VII. CONCLUSION

Although the participant pool must be increased in order to make any firm quantitative claims, this study adds to the conversation of CALL for indigenous and endangered languages. The perceptions of the users have provided us with an extensive list of changes to be considered for *nêhiyawêtan* that has been heuristically evaluated (see Appendix D).

For CALL for CILs generally, it is important to consider the layout of the interface. Are the instructions visible? Do they fall within the users' natural line of sight? The design of the layout should also be culturally relevant, and respectful of any social constraints of the specific language community the program is developed for.

Another aspect that will be university course/community dependent is the dialect that is chosen. If the program is intended for broad use it would make sense to use a standard or prestige dialect. However, if it is made specifically for a certain course or community, the program should reflect the dialect being used most within the course/community. Failure to consider the dialect of the unique target group could actually have a detrimental effect on the prestige of the local/focus dialect. If the course utilizes non-standard conventions, the pros and cons should be weighed in order to ensure the language is taught efficiently and in a way that won't be detrimental to the use of the language. Overall, the creation of an ICALL program should have a positive effect on the prestige of the CIL in question.

The program should not require the user to leave the interface for help. The participants in this study were less likely to use a function that required leaving the interface than one that was built into the interface. If it is possible to embed the function within the interface, it would be more user-friendly to incorporate the function in that manner. *nêhiyawêtan* in particular, has already shown that the interface is capable of embedding this function and the inconsistency of its implementation is puzzling to users.

The group undertaking CALL program development must also consider the vocabulary resources. There must be a large amount of vocabulary to avoid repeating questions. This will also allow the user to practice for longer before they exhaust the exercise questions.

Additionally, although audio was not a main focus for *nêhiyawêtan* in particular, it is a feature that should be incorporated into CALL programs for CILs. Even though *nêhiyawêtan* is meant to create the opportunity to practice speaking the language during class time by providing a grammatical resource outside of class time, audio should still be incorporated into the application. By linking audio to the grammatical paradigm practice, students will learn the pronunciations and be better able to understand other speakers. This will give students a better idea of the orthographical representation of the phonemes of the language. After all, the end goal for both methods of language learning is for the learner to be competent in all aspects of the language.

It is likely that this ICALL program is the first of its kind focusing on a CIL. This progress made for CALL for CILs is important to expand the domain of the language. By encouraging learners and speakers to engage with their language online, it increases its use and accessibility, especially in today's technology-dominated society.

In the most basic sense, ICALL presents CILs with a learning tool that had not existed previously. This advancement allows for more flexibility among CIL instruction techniques and provides instructors with an updated way of teaching their materials. The feedback learners can now access

outside of class time has the potential to allow instructors to completely reassign in-class time to other aspects of the language that had been limited or excluded previously. For example, teachers may choose to spend the majority of in-class time practicing conversational aspects of the language since the learners have other speakers present to converse with, which they most likely do not have outside of class time. Many language classes spend the majority of time learning vocabulary and drilling paradigms; ICALL creates the opportunity for learners to learn these aspects outside of class-time, and giving them more opportunity for speaking practice which is harder to obtain outside of the classroom.

Overall, practicing morphology outside of class time is not only possible, but useful as well. User evaluations were generally positive and the idea of ICALL for a CIL was well received. It is hoped that the experiences of these participants will help others to design their own CALL projects for indigenous languages. The feedback that they gave with regards to the interface and design is something that can be perceived independent of the language, and will easily apply to other polysynthetic languages. Future studies should include a greater number of participants so that the data may be quantified. University projects should be compared to those stemming from the communities directly, and collaborations should be made to ensure the programs account for diverse user populations. There is still much to be evaluated when creating CALL programs for CILs, and many contexts must be analysed in order to gain a wider perspective on the matter.

## REFERENCES

- ÅKERSTEDT, ULF. 1999. *Mätbar användbarhet*. Stockholm: Nada, Kungliga Tekniska Högskolan. Description: Practical technique for usability studies.
- ANTONSEN, LENE. 2013. Constraints in free-input question-answering drills. Proceedings of the second workshop on NLP for computer-assisted language learning at NODALIDA 2013. *NEALT Proceedings Series 17 / Linköping Electronic Conference Proceedings* 86.11–26.
- ANTONSEN, LENE; RYAN JOHNSON; TROND TROSTERUD; and HELI UIBO. 2013. Generating modular grammar exercises with finite-state transducers. Proceedings of the second workshop on NLP for computer-assisted language learning at NODALIDA 2013. *NEALT Proceedings Series 17 / Linköping Electronic Conference Proceedings* 86.27–38.
- BELL, DOUGLAS S.; CHARLES E. HARLESS; JERILYN K. HIGA; ELIZABETH L. BJORK; ROBERT A. BJORK; MOHSEN BAZARGAN; and CAROL M. MANGIONE. 2008. Knowledge retention after an online tutorial: A randomized educational experiment among resident physicians. *Journal of General Internal Medicine*, 23.1164-1171
- DEHCHO FRIENDSHIP CENTER. 2015. Dene K'ęę Gúdeh. Online: <http://gudeh.com/#>.
- CARROLL, SUSANNE, and MERRILL SWAIN. 1993. Explicit and implicit negative feedback. *Studies in Second Language Acquisition* 15.357-386.
- CHAPELLE, CAROL A. 2001. Computer applications in second language acquisition: Foundations for teaching, testing and research. *Cambridge: Cambridge University Press*.
- COBB, TOM, and VANCE STEVENS. A principled consideration of computers & reading in a second language. *The Power of CALL*. 115-136.
- FELIX, USCHI. 2003. Language learning online: Towards best practice. *Lisse: Swets & Zeitlinger*.
- FIRST VOICES. 2013. Apps. Online: <http://www.firstvoices.com/en/apps>.
- GARRETT, NINA. 1987. A psycholinguistic perspective on grammar and CALL. *Modern Media in Foreign Language Education: Theory and Implementation*. Lincolnwood, IL: National Textbook. 169-96.
- HAGDAHL, KARL-HENRIK. 2000. *Checking the checker grammatifix: analysis of the Swedish grammar checker from a user's point of view*. Royal Institute of Technology: Master's Thesis in Computer Science at the School of Computer Science and Engineering.
- HARVEY, DAN. 2015. ACORNS overview. University of Southern Oregon, OR. Online: <http://cs.sou.edu/~harveyd/acorns/index.php>
- HEFT, TRUDE. 2001. Error-specific and individualised feedback in a web-based language tutoring system: Do they read it?. *ReCALL* 13.99-109.

- HEIFT, TRUDE. 2003. Multiple learner errors and meaningful feedback: a challenge for ICALL systems. *CALICO Journal* 20.533-548.
- HEIFT, TRUDE, and ANNE RIMROTT. 2012. Task-related variation in computer-assisted language learning. *The Modern Language Journal* 96.525-543.
- HO CHUNK NATION. 2015. Ho Chunk. Android Application.
- JUNKER, MARIE-ODILE, and DELASIE TORKORNOO. 2012. Online language games for endangered languages. *Proceedings of EDULEARN 12: International Conference on Education and New Language Learning Technologies*. Barcelona, Spain.
- MARIAN, VICTORIA; HENRIKE K. BLUMENFELD; and MARGARITA KAUSHANSKAYA. 2007. The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research* 50.940-967.
- MONK, ANDREW; PETER WRIGHT; JEANNE HABER; and LORA DAVENPORT. 1993. Improving your human-computer interface: a practical technique. *BCS Practitioner Series*. 1-11.
- NAGATA, NORIKO. 1993. Intelligent computer feedback for second language instruction. *The Modern Language Journal* 77.330-339.
- NAGATA, NORIKO, and M. VIRGINIA SWISHER. 1995. A study of consciousness-raising by computer: The effect of metalinguistic feedback on SLA. *Foreign Language Annals* 28.336-347.
- NAVAJO LANGUAGE RENAISSANCE. 2008. Free demo. Online: <http://navajorenaissance.org/demo.html>.
- NESBITT, DALLAS. 2013. Student evaluation of CALL tools during the design process. *Computer Assisted Language Learning* 26.371-387.
- PIIKANI SOCIETIES. 2015. Nitsi Poh Sin. Android Application.
- ROBERTS, ALICIA. 2015. Southern Tlingit 1. Android Application.
- ROSETTA STONE. 2015. Endangered languages. Online: <http://www.rosettastone.com/endangered/projects>.
- SIMON FRASER UNIVERSITY. 2015. Haida language. Online: <http://haida.intelligentlanguagetutor.com/>.
- SON, JEONG-BAE. 2007. Learner experiences in web-based language learning. *Computer Assisted Language Learning* 20.21-36.
- STRUTHERS, ROXANNE, and CYNTHIA PEDEN-McALPINE. 2005. Phenomenological research among Canadian and United States indigenous populations: oral tradition and quintessence of time. *Qualitative Health Research* 15.1264-1276.

- SOVEREIGN ONEIDA NATION OF WISCONSIN. 2015. Oneida Language Application. Android Application.
- UiT THE ARCTIC UNIVERSITY OF NORWAY AND UNIVERSITY OF ALBERTA. 2015. *nēhiyawêtan*. Online: <http://oahpa.no/neh yawetan/>.
- WARD, MONICA, and JOSEF GENABITH. 2003. CALL for endangered languages: challenges and rewards. *Computer Assisted Language Learning* 16.233-258.
- WOLVENGREY, AROK. 2001. *nēhiyawēwin itwēwina*. University of Regina: Canadian Plains Research Center.

APPENDIX A  
A Design Study

### *Procedure*

The design study had been completed prior to the start of this study. The idea was to familiarize the researcher with the application so that they would be able to anticipate the feelings and experiences of the users. The design study enables the researcher to create a more relevant questionnaire and ask the user appropriate questions during each session.

The design study followed the same procedure, however, the location was a home-office, and a camera was used for the recording instead of screen-casting. There were also some changes to the tasks, although the time frame was similar. The tasks included Leksa (Cree-to-English, English-to-Cree), Morfa-C verbs (past and present tense), Morfa-S nouns (plural, locative, and diminutive), Morfa-S verbs (past and present).

### *Results*

Question Type	Questions Set Scores (Of 33 Sets)		
	<sup>a</sup> M	SD	<sup>b</sup> Range
Morfa-S Plural	3	1	2-4
Morfa-S Diminutive	3	1	2-3
Morfa-S Locative	4	1	3-5
Morfa-S Present	3	0	3-3
Morfa-S Past	1	2	0-4
Morfa-C Present	5	1	4-5
Morfa-C Past	3	3	0-5
Leksa Cree-to-English	3	1	2-4
Leksa English-to-Cree	2	2	1-4

<sup>a</sup>M: mean. <sup>b</sup>Range: 0 to 5 (full marks).

Table A1.1: *Scores for each exercise type.*

Like in the user evaluation study, the scores for Morfa-S past tense exercises were the lowest. This shed light on the potential problem and encouraged a more in-depth evaluation of this particular exercise in the current study.

% of all sets (of 33 sets)	
User showed answers	
When all answers were 'correct'	42%
After some self-correcting	52%
User did not show answers	
All answers 'were correct'	6%

Table A1.2: *Percentages of the number of question sets during which the 'show answers' button was used or ignored.*

The higher rate in comparison to the user evaluation study is most likely a result of the researcher wanting to thoroughly test the functions of the application.

% of all times help was available (18 times total)	
Used help	
Helpful	28%
Not helpful (does not address the problem)	39%
Did not use help	
Would not have been helpful	33%

Table A1.3: *Percentages of the number of times the 'help' function was available, whether or not it was used and whether or not it was helpful.*

In the design study, it was concerning that over 70% of the time, the 'help' function was off-topic. This inspired a more in-depth error analysis in the following study to observe a greater variety of potential user errors.

The most common reasons for lack of helpfulness were:

- Feedback did not address vowel length errors.
- The feedback was unavailable for a specific word.
- The computational model was incorrect.

The final two reasons were remedied prior to the user evaluation.

Extent of Self-Correcting	% of initial errors (38 errors)
User completely corrected answers	66%
User attempted to correct answers	29%
User left the question blank.	5%

Table A1.4: *Percentage of the number of times self-correcting occurred and to what extent.*

It is most likely that answers are completely corrected in the design study more often than in the user evaluation because the researcher was more interested in possible user pathways to achieve the correct answer. Whereas, the average user would not have any interest in various patterns of user thinking.

### *Heuristic Evaluation*

Severity ratings were assigned to each problem and usability heuristics were addressed during each session (Hagdahl 2000):

- **Very High**- The problem must be fixed in order to maintain user trust of the program. Will deter all use of the program.
- **High**- The problem needs to be fixed since it is an inconvenience for the user.
- **Medium**- The problem should be eventually be fixed but is not a great concern for program use.
- **Low**- The problem should be looked at and eventually replaced by a more efficient design.

Below are the severity ratings of various problems encountered:

- On average, the ‘test answers’ function was used 2-3 times per set for verb exercises, and just 1-2 times per set for noun exercises. Leksa had the least amount of testing of all the exercises. Undoubtedly, these results are due to the rigid structure of most vocabulary exercises.

**Low** severity: The noun and verb testing is proportionate to the complexity of the exercise. However, it would be helpful to create user feedback for Leksa to encourage user re-testing.

- The ‘help’ function was used 67% of the times that it became available. However, 72% of the times the messages did not address the error. Vowel length errors were the most common, while the help only addresses errors of inflection.

**High** severity: The “help” function is often unhelpful and this could be frustrating to users. One of the main goals of this application is to provide users with immediate feedback and it is important that it is reliable.

- Feedback was unavailable for certain words and exercises.

**Very High** severity: The “help” function is inconsistent and this could be frustrating to users. One of the main goals of this application is to provide users with immediate feedback and it is important that it is reliable.

- The computational model was incorrect.

**Very High** severity: If the computational model is incorrect, the application could be teaching students incorrect forms. It is essential that the computational model is corrected.

- Approximately 41% of the vowel length errors were marked as correct, whereas 59% were marked as incorrect. The answers marked as correct contained either one vowel length error, or multiple errors concerning the same vowel.

**Very High** severity: Users are given inconsistent feedback and no warning or explanation of this inconsistency.

These heuristics provided a basis for the user evaluation study, to reinforce which problems were truly pervasive.

APPENDIX B  
Study Checklist

Practical Details	Intention	Motivation
What tasks should be completed during each session?	<p>Morfa-S: plural and diminutive nouns, past tense verbs</p> <p>Morfa-C: plural, locative and possessive nouns, present tense verbs</p> <p>Leksa: Cree to English and English to Cree vocabulary</p>	<p>This will expose the user to all the possible functions of the application: e.g. the “help” functions for each of these exercises, the “click on a word” function, audio files, etc.</p> <p>The varied level of difficulty of these exercises should keep the user from being overwhelmed or bored.</p>
Should users test multiple applications and compare them?	No	This would take too much time. This will be briefly addressed by the questionnaire issued to the participants at the end of each session.
How many participants should evaluate the application?	6	Hagdahl’s (2000:33) study opted for 6. This is a fitting number because the goals of these two study coincide. This should be enough users to reveal major application flaws. This is also suited to the relatively restricted population size of the participant pool.
How will the users be introduced to the application, and for how long?	5 minutes, maximum	<p>I would like users to explore the interface in a way they would working alone at home.</p> <p>I simply tell the users the information they need to navigate to a particular exercise: i.e. “We will now work in Morfa-S on plural nouns.”</p> <p>I also make sure users know the keyboard shortcut for the vowel length markers.</p> <p>Aside from this brief instruction, I encourage the user to rely on the given instructions to complete the practice exercises.</p>

		By doing this I hope to capture beginners' errors as well as uncertainty of instructions. This should also capture the user strategy of the participant, and in what manner they use the interface (and whether this manner is the intended manner of use).
What background/education should the participants have?	Some NS 152 or higher	<p>The application was developed using the NS 152 course material. It is important to obtain the opinions of students at the beginner level. This is estimated to provide feedback on the help and instructions.</p> <p>Participants that have taken upper level Cree courses will also have useful input. Since they will be more likely familiar with the material, they should have less material-based problems and be able to focus more on the interface itself.</p>
Should the participants work alone or in pairs?	Participants should work alone.	The participants should work alone, since this will most likely be the environment they would use this application in. The application is meant to provide the user with feedback on their answers when no other help is available for immediate feedback.
Should participants be instructed and prompted to think out loud?	Yes, to a certain degree.	First, participants will be instructed to think out loud (e.g. make any difficulties, frustration, elation, or observations clear). However, the user will have more control over the amount of engaging than in Haggdahl (2000:33). These tasks require a greater degree of user concentration since the participant is already required to produce inflected forms in an L2 or L3. Due to this mentally taxing task, the user was only asked to explain when they initiated, or emoted.
Should participants be encouraged to ask questions?	Yes, to a certain degree.	The participants should be encouraged to ask questions because it gives insight into user problems. However, students will be discouraged from asking questions without first reading the instructions or attempting to figure things out by themselves. This is important because the participant would

		realistically be relying solely on the instructions when using the application outside of class time.
How should the researcher respond to participant questions?	Dependent on the questions and their context.	The researcher will give limited information regarding help with the material (the help button should be consulted first), and the question format (the instructions should be consulted first). This will help to uncover which instructions, and feedback are actually unclear.
Should the researcher act as a help system for the user?	No	There is a help system available, and it is important that it is tested.
Where/in what environment should the session take place?	Laboratory/ Controlled quiet room.	All sessions will take place in the phonetics lab in a sound proof booth. If this booth is not available another quiet room will be sought out. This will allow for a relatively controlled environment.
How long should each session be?	1 hour maximum	Motivation and focus of the participants are expected to drop after a longer period of time, especially when exposed to repetition-based grammatical exercises.

Table B1.1: *Complete Study Checklist*

APPENDIX C  
User Evaluation

## User Evaluation Survey

Age: \_\_\_\_\_ Gender: \_\_\_\_\_ First Language: \_\_\_\_\_

Other languages studied: \_\_\_\_\_

Where did you learn Cree? If at the university, list which courses you took.

How would you rate your knowledge of Plains Cree:

Reading:

Good	Fair	Poor
------	------	------

Writing:

Good	Fair	Poor
------	------	------

Listening:

Good	Fair	Poor
------	------	------

Speaking:

Good	Fair	Poor
------	------	------

How would you rate your computer skills?

Good	Fair	Poor
------	------	------

Have you used a computer-aided language learning program before? If so, list the programs you have used and write a brief description of what you did or did not like about them.

Do you think this program complements the Cree course you took? If yes, why so? If not, what would you want the program to offer?

Does the design of this program make the user learning goals clear? Do these goals coincide with your learning goals?

When given contextual questions do you read the entire question? Or focus on the information necessary to fill in the blank?

Did you notice the “click on a word” function? Was it helpful?

Did you notice the “help” function? Was it helpful?

What did you like most about the demo?

What did you like least about the demo?

What do you think most needs changing?

Did you find the recording equipment intrusive?

Any problems you would want to address?

## APPENDIX D

### Suggested Improvements & Heuristic Evaluation

The same heuristic evaluation used in the design study is used to evaluate the problems found during the user evaluation study. A description of the severity levels can be found in Appendix A.

- Leksa had the least amount of testing and the most answers left blank out of all the exercises.

**Low** severity: The noun and verb testing is proportionate to the complexity of the exercise. However, it would be helpful to create user feedback for Leksa to encourage user re-testing. This is consistent with the results from the design study.

- Users had trouble finding instructions and parts of speech (POS) options.

**High** severity: Relocating the instructions and POS options is a simple fix that will have a big impact on the usability of the application. Currently users are ignoring the instructions because they are not located in an expected location. The instructions should be moved directly above the question sets. The current isolation of the POS options is unnecessary. The links to change the POS would be more noticeable if grouped with the rest of the exercise type options.

- ‘Click on a word’ is only available for certain words.

**Medium** severity: The function is currently working properly, however it needs to be implemented for more words. There are currently other resources for users to utilize if they want the translation of the word; although, ideally the ‘click on a word’ function will be extended to incorporate more words.

- Audio incorporation is lacking.

**Low** severity: Though it would be ideal to have audio for each question, it is not a current priority of this application.

- Users didn’t read the context of Morfa-C questions.

**High** severity: The goal of Morfa-C is to allow users to practice inflection within the context of a sentence. If the sentence itself is not required to input the answer, the question is improperly constructed and should be remodeled to provide the user with a more engaging context.

- Users found the formatting of Morfa-S verb questions to be confusing.

**High** severity: This exercise had the lowest ranging scores because of its somewhat ambiguous question format. The exercise should be re-modeled and re-evaluated.

- The content was not consistent with the textbook.

**High** severity: It is important for beginners to have reliable and consistent feedback. The application should adhere to the conventions of one resource/dialect/orthography in order to avoid confusing learners.

- User would like to review the paradigm before starting the exercises.

**Low** severity: Though it would be nice to have this option, with the current inconsistencies it would be better for the user to confide in their textbook before practicing with the application.

- Users would like updated interface design.

**Medium** severity: Users were not particularly put-off by the current design. It is most important that the application is running properly. However, it is important to develop a culturally relevant design and appropriate exercise names.