

Creating an Online Hul'q'umi'num' Dictionary for Teachers and Learners

by

Helen Yu Zhang

B.A., University of Massachusetts Amherst, 2017

B.A., University of Rhode Island, 2009

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

© **Helen Yu Zhang 2023**
SIMON FRASER UNIVERSITY
Spring 2023

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name: Helen Yu Zhang
Degree: Master of Arts
Thesis title: Creating an Online Hul'q'umi'num'
Dictionary for Teachers and Learners
Committee: **Chair:** Heather Bliss
Lecturer, Linguistics

Nancy Hedberg
Supervisor
Professor, Linguistics

Donna B. Gerds
Committee Member
Professor, Linguistics

David Beck
Examiner
Professor
Department of Linguistics
University of Alberta

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

This thesis is about a collaborative dictionary project for Hul'q'umi'num', the Island dialect of Halkomelem, a Salish language spoken in British Columbia, Canada. The goal of the project was to create a teachers' and learners' resource with words and sound, organized by categories designed to be useful for the preparation of language materials. I worked alongside language instructors, students, and linguists to organize legacy materials and collect new media to build an accessible, sustainable dictionary database and to create a learners' dictionary website for public viewing.

Keywords: Halkomelem; Salish; dictionary; digital lexicography

Dedication

in memory of my njabu (grandma)

&

a special shoutout to my cat

Acknowledgements

hay tseep q'a' all of the hul'q'umi'num' and hənqəminəmə students and teachers.

hay tseep q'a' all of the elders who have shared and continue to share their knowledge, words, and voices for the future.

i am especially grateful to the late elder dr. ruby peter, *sti'tum'at*, and the elder delores louie, *swastanulwut*, for all of their work and dedication to the hul'q'umi'num' language, and for sharing some of their time with me.

—

hay ch q'a' nancy hedberg, for helping me navigate the world of academia.

hay ch q'a' donna gerdts, for bringing me into the world of language revitalization.

—

some colleagues, in the order that we met — zack gilkinson, maida percival, lauren schneider, michelle parent, tess nolan, chloë farr, and rosemary webb. thank you to zack for assisting me with typesetting \LaTeX .

sharon seymour, evangeline paige, thomas johnny, b. ruby peter, and all the members of the hul'q'umi'num' language & culture society

abigail cunningham, kaitlyn cunningham, and all the members of the x^wəlməx^wqəən cultural society

—

the following people, who had helpful conversations with me — janet leonard, john lyon, david beck, and charles ulrich.

thomas hukari

—

my irl friends elise, titilayo, and srijani

maria, who i'm forcing to be my next irl friend

—

my mom

the following organizations have provided various funding opportunities for me to do this work - the hul'q'umi'num' language & culture society, the first peoples' cultural council, the simon fraser university department of linguistics, and the social sciences and humanities research council of canada.

Table of Contents

Declaration of Committee	ii
Ethics Statement	iii
Abstract	iv
Dedication	v
Acknowledgements	vi
Table of Contents	viii
List of Tables	x
List of Figures	xi
Preface	xii
1 Introduction	1
1.1 Hul'q'umi'num' and language documentation	4
1.2 Dictionaries in Hul'q'umi'num'	6
1.3 Roadmap	8
2 A tour of the dictionary	9
2.1 Components of the website	10
2.1.1 Topics	10
2.1.2 File card entries	11
2.1.3 Roots	12
2.1.4 Lexical suffixes	13
2.1.5 Audio	15
2.1.6 Images	15
2.2 Dictionary structure	16
2.2.1 Front matter - about the dictionary	16
2.2.2 Microstructure - the file card entry	16

2.2.3	Macrostructure - arrangement of words	17
2.2.4	Cross-reference structure - linking words	17
2.2.5	Access structure	19
2.2.6	Other	19
3	Building the dictionary	20
3.1	The initial data situation	22
3.2	Text processing	24
3.3	From spreadsheet to database	27
3.4	From database to website	32
3.4.1	Website design	34
3.4.2	Alphabetical ordering	35
4	Conclusion	37
4.1	Reflecting back	37
4.1.1	A note on collaborative and participatory models of research	37
4.1.2	A note on project management	38
4.2	Looking ahead	39
4.2.1	More dictionary content	39
4.2.2	Capacity building	39
4.2.3	Re-use and extensibility	40
4.2.4	Future data migration	40
4.3	Final thoughts	41
	Bibliography	43
	Appendix A Hul'q'umi'num' Practical Orthography	50
	Appendix B Hul'q'umi'num' Sound Inventory	51
	Appendix C Full List of Dictionary Topics	54
	Appendix D Transliterating between APA and the Practical Orthography	58
	Appendix E Transcoding legacy fonts - Straight	61

List of Tables

Table 2.1	Words with = <i>ew't-hw</i>	15
Table 2.2	Some food related words	18
Table 3.1	WordPress to Dictionary mapping	34

List of Figures

Figure 1.1	Map: Halkomelem	5
Figure 1.2	Map: Island Hul'q'umi'num'	6
Figure 2.1	Home page	9
Figure 2.2	Website - Topics	10
Figure 2.3	Website - Verb roots	12
Figure 2.4	Website - Verb root example	13
Figure 2.5	Website - Lexical suffixes	14
Figure 2.6	Website - Lexical suffix example	14
Figure 2.7	Website - Detailed entry	17
Figure 3.1	Incorporating new and legacy work	24
Figure 3.2	Lexical suffix workflow part 1	31
Figure 3.3	Lexical suffix workflow part 2	31
Figure 3.4	Playing audio	32
Figure 3.5	Results for “bark”	33
Figure 3.6	Dashboard	33
Figure 3.7	User groups	34
Figure 3.8	Website - Index	36
Figure 3.9	Alphabetical ordering of the first sound	36

Preface

My first visit to Duncan, BC, took place in the spring of 2019, and that was also the beginning of my involvement with the Hul'q'umi'num' Language & Culture Society (HLCS). The moment I stepped into the *Shhwulmuhwqun* in Duncan I was introduced into a fast-paced environment, with students full of passion and stories to tell. The initial planning for the dictionary was done in conjunction with a course that took place later that summer for a mixed cohort of BA and MA students taught by Professor David Beck. We were assisted by elders Delores Louie and Ruby Peter and supervised by linguist Donna Gerdts. The original plan was essentially a word list with sound files, totalling perhaps 1,000 words for a simple “children’s dictionary”. This idea quickly grew into something else.

I made regular trips between Vancouver and Vancouver Island until early 2020, when I returned to Vancouver on a ferry trip dated February 29th. Days later, work on the dictionary ground to a complete halt along with the rest of the world. Previously, all classes and language revitalization activities had taken place on Vancouver Island, but I would not set foot again on the island until summer of 2022. I spent the rest of 2020 trying to help navigate the sudden transition to remote learning by digitizing language resources for the website and researching Zoom-friendly activities for the language classes.

At the start of 2021, I was introduced to a small team of linguists and students who were in the process of digitizing and developing materials for *hənqəmihəm*, the Downriver dialect. I spent most of 2021 with these projects, which, while not the focus of this thesis, have greatly affected my thinking about the Island dictionary project.

It was only in 2022 when I was able to return my focus back to the original dictionary project.

Chapter 1

Introduction

The province of British Columbia is home to 60% of Canada's Indigenous languages. Almost all of these are considered severely endangered, with fluent speakers making up less than 3% of the Indigenous population (Gessner et al., 2022). This project focuses on the language of the Hul'q'umi'num', a Coast Salish people who live along the western shores of the Salish Sea. Today only around forty fluent first-language speakers remain, mostly over the age of seventy-five (D. Gerds, p.c.). However, the use of the language remains strong in traditional cultural practices, in some homes, and in school programs including pre-school language nests, K–12 classrooms, and post-secondary language courses. Hul'q'umi'num' is one of three dialects of Halkomelem.¹ Currently there are 1,900 active L2 learners across the three dialects combined (Hul'q'umi'num', həŋqəmihəm, and Halqeméylem), including an estimated 105 fluent and 512 semi-fluent speakers (Gessner et al., 2022). Many of the learners are highly motivated to become proficient speakers and champion the language as teachers, researchers, material developers, and language revitalizers. The new speakers see it as their responsibility to become fluent in the language so that they can transmit it to the younger generations in their communities and families (Gessner et al., 2022; Luning and Yamauchi, 2010). It is with this goal in mind that Indigenous and non-Indigenous scholars involved in language documentation give top priority to creating resources for use by those seeking to strengthen the language.

¹Halkomelem is an anglicization of the name of the Upriver dialect and has been used to refer to the language as whole.

For the past four years, I have been working alongside the instructors and students in the Simon Fraser University Hul'q'umi'num' language program.² This cohort-based program offers certificates and diplomas in Indigenous Language Proficiency, BA degrees (with minors in Indigenous Languages and Linguistics), and MA degrees in Indigenous Languages and Linguistics. During this time, I have provided technical support for SSHRC-funded projects on legacy and new Hul'q'umi'num' materials. Some of the things I worked on include creating text processing tools for converting between orthographies, digitizing legacy materials,³ and developing interactive website resources to support remote learning. This work led to my involvement with the Hul'q'umi'num' Language & Culture Society, a team of Elders, language champions, and academic partners, who works collectively to provide community-based and university Hul'q'umi'num' language programs and projects. In 2019, I began working with them on an online dictionary project, which has been funded in part by a Digitization grant from the First Peoples' Cultural Council. This thesis details this project, whose goal was to create a teachers' and learners' resource with words and sound, organized by categories designed to be useful for the preparation of language materials.

Dictionaries have always been an important part of the language documentation process (Himmelman, 1998). Academic dictionaries of Indigenous languages are comprehensive and dense out of necessity, a format that is not ideal for language learning which “imposes its own requirements on the format of a dictionary” (Svensén, 1993). Despite the immediate needs of critically endangered languages, compiling a proper academic dictionary may take years or even a lifetime to compile (Hargus, 2019; Rood and Koontz, 2002). As part of traditional documentary linguistics, academic dictionaries can funnel language into a “one-way journey,” in which “language speakers are isolated from representations of language events, the knowledge encoded in them, and the hard-won linguistic explanations and outcomes” (Nathan and Fang, 2013, p. 43). In recent years, attention has turned to how these dictionaries can play a more active

²SFU INLP - <https://www.sfu.ca/inlp.html>

³This includes digital versions of Gerdt (1996, 1997) and Gerdt et al. (1998).

role in language revitalization (Berez-Kroeker et al., 2021; Callaghan, 2002; Hinton and Weigel, 2002; Lew, 2014; Weber, 2021; Yong and Peng, 2007). Thematic “mini” dictionaries can be small projects done with limited resources that immediately go to the community (Cablitiz, 2011; Mosel, 2011), while a hybrid approach to dictionary making involves prioritizing the faster creation of a learners’ dictionary while a comprehensive dictionary is slowly developed in the background (Stebbins, 2003).

While early dictionaries were limited by technology,⁴ the internet, along with modern computational power, has revolutionized the lexicography field (Dziemianko, 2017; Nielsen and Tarp, 2009; de Schryver, 2003; Tarp, 2019). New tools make it easier to organize lexicography work, and modern website publishing platforms such as WordPress feature visual editors that make it easy to publish content. Access to internet service in most locations in the Hul’q’umi’num’ territory and the widespread use of computers and smartphones make online materials available and useful to the language community. For these and other reasons (cf. Zhang et al., 2021), creating an online dictionary for Hul’q’umi’num’ seemed like an excellent idea to me.

The dictionary team consisted of various members of the Hul’q’umi’num’ community including Elders, undergraduates, and graduate students, as well as Professor Donna Gerdts. The goals of the project were threefold — to build an dictionary database, to create a learners’ dictionary website for public viewing, and to design a sustainable workflow connecting the database and website. The desired features were:

- A dictionary website viewable on small devices (such as phones or tablets)
- A dictionary website where learners could listen to audio files of the words
- A dictionary website that would be easy to navigate for community members who might not be tech savvy
- A dictionary website that would be interesting and informative for users who may not have an academic background in language structure

⁴*Making Dictionaries: Preserving Indigenous Languages of the Americas* (Frawley et al., 2002) collects numerous examples of the challenges that earlier linguists faced.

- A dictionary website that would be useful for users with an academic background in linguistics
- A dictionary website and database that can be easily and frequently updated by language team (e.g. a living dictionary)
- A database that could be worked on by people with Mac or Windows computers
- A database where the data can be exported in a non-proprietary format

The remainder of this introduction gives some further background information that helps situate this project. Section 1.1 introduces the Hul'q'umi'num' language and its documentation and Section 1.2 gives a brief survey of various Hul'q'umi'num' dictionaries.

1.1 Hul'q'umi'num' and language documentation

Halkomelem is one of twenty-three Salish languages traditionally spoken in southwestern British Columbia and the northwestern United States. Halkomelem, a Salish language, was historically spoken on Vancouver Island and in the Lower Mainland from Tsawwassen up the Fraser Valley to Yale. Halkomelem is comprised of three dialect groups: Hul'q'umi'num', the Island dialect spoken on Vancouver Island; həŋqəmihəm, the Downriver dialect spoken on the lower mainland near the mouth of the Fraser River (Suttles, 1990); and Halqeméylem, the Upriver dialect spoken further up the Fraser Valley. The map⁵ in Figure 1.1 shows the approximate area where Halkomelem is traditionally spoken.

The Hul'q'umi'num' dialect is spoken by various First Nations and Bands in southwestern Vancouver Island and neighboring islands. Please see Figure 1.2 for peoples in the Hul'q'umi'num' territory.

There has been much research and documentation done on Hul'q'umi'num' over the past few decades. Thanks to the speakers that have aided linguists in their research over

⁵Both maps in this thesis were drawn using the Felt mapping tool (<http://felt.com>) and the watercolor map tiles by Stamen Design (<http://maps.stamen.com/>).



Figure 1.1: Halkomelem

the last 130 years, much is known about the linguistic structure of Hul'q'umi'num' (Gerdts, 1988, 2003, 2010; Gerdts and Hinkson, 2004; Gerdts and Hukari, n.d., 2000, 2004, 2006, 2011; Gerdts and Werle, 2014; Hukari, 1978, 1981; Leslie, 1979b; Rozen, 1985; Urbanczyk, 2004).

In recent years, Hul'q'umi'num' students in the Hul'q'umi'num' Language Program have completed master's projects on the language on diverse topics such as place names (Seward-Wilson, 2019), kinship terms (Charlie, 2019), birds (Sylvester, 2019), traditional art (Marston, 2021), storytelling performance (Claxton, 2020) and ceremonial wedding language (Sam, 2019). These projects carefully situate the language in its cultural context.

Hul'q'umi'num' is a challenging language for speakers of English to learn due to its complex phonetic system.⁶ Hul'q'umi'num' has 37 consonants, 24 of which do not occur in English. Hul'q'umi'num' has ten back-of-the-mouth sounds /k k^w ḳ^w q q̣ q̣^w x̣, x̣^w,

⁶See <http://sqwal.hwulmuhwqun.ca/learn/learning-the-sounds/alphabet/>



Figure 1.2: Island Hul'q'umi'num'

a complex set of coronal affricates /t^h t^h' c č č č'/, three lateral sounds /l ł ʎ/. It is also unusual in having plain laterals /m n l y w/ contrasting with glottalized ones /m̥ n̥ l̥ y̥ w̥/.

Over the years, several orthographies have been developed for Hul'q'umi'num'. The practical orthography used in this thesis was adopted in 2002 by the Hul'q'umi'num' Treaty Group elders' committee associated with a SSHRC CURA grant and the University of Victoria. The materials developed by the Hul'q'umi'num' Language & Culture Society use this orthography.⁷

1.2 Dictionaries in Hul'q'umi'num'

There has been much research and documentation done on Hul'q'umi'num' over the past few decades. This includes several Hul'q'umi'num'-English dictionaries that have been published over the years.

- **1971 Nanaimo Dictionary** (Leslie, 1979a)

⁷See Appendix A for a table listing the characters of the practical orthography and their APA equivalents.

Adrian Leslie, a University of Victoria student, compiled a list of 1,400 words arranged semantically based on his fieldnotes. They are written in Dr. O'Grady's Cowichan alphabet.

- **1978 Nanaimo Dictionary** (White, 1978)

The late Ellen White of Snuneymuxw compiled and published a 60-page dictionary.

- **1995 The Cowichan Dictionary of the Hul'q'umi'num' Dialect of the Coast Salish People** (Hukari and Peter, 1995)

This 6,600-word dictionary, edited by Tom Hukari and the late Dr. Ruby Peter compiles work done with the Cowichan elders and is in the older Cowichan Tribes orthography. The words are presented in 122 pages of Hul'q'umi'num' to English followed by 212 pages of English to Hul'q'umi'num', which also include example sentences for some words. The dictionary also includes a 35-page introductory grammar.

- **1997 Hul'q'umi'num' Words: An English-to-Hul'q'umi'num' and Hul'q'umi'num'-to-English Dictionary** (Gerdt et al., 1997)

This 263-page dictionary was published in the Cowichan Tribe orthography. It was designed for classroom use and an effort was made to include everyday vocabulary. The first half of the dictionary contains nouns organized by semantic topic. Ethnobotany and ethnozoology information, as well as some cultural notes are included. The dictionary contains a bi-directional index with Hul'q'umi'num' to English and English to Hul'q'umi'num'. There are 2,142 nouns and 1,300 verbs.

- **1999 Halkomelem Talking Dictionary** (Gerdt and Penrowley, 1999)

This early "talking dictionary" was distributed on a CD and included 3,500 sound files recorded by Elders of the words in the 1997 *Hul'q'umi'num' Words* dictionary. The talking dictionary was created using ToolBook, software which is now defunct.⁸

⁸<https://www.sumtotalsystems.com/toolbook>

- **2007 Quw'utsun Hul'q'umi'num' Category Dictionary** (Cowichan Tribes, 2007)
This 120-page dictionary consists of approximately 3,000 words arranged by semantic category.

1.3 Roadmap

This thesis is intended to serve as a practical guide to the Hul'q'umi'num' online dictionary, both for learners who will use it and for anyone who may find themselves needing to build such a resource. Chapter 2 will function as a tour of the website and dictionary structure. Chapter 3 will explain the technical aspects of what was involved in creating the dictionary. Chapter 4 will serve as a conclusion, with some final thoughts on how this iteration of a community dictionary came together the way it did and prospects for the future.

Chapter 2

A tour of the dictionary

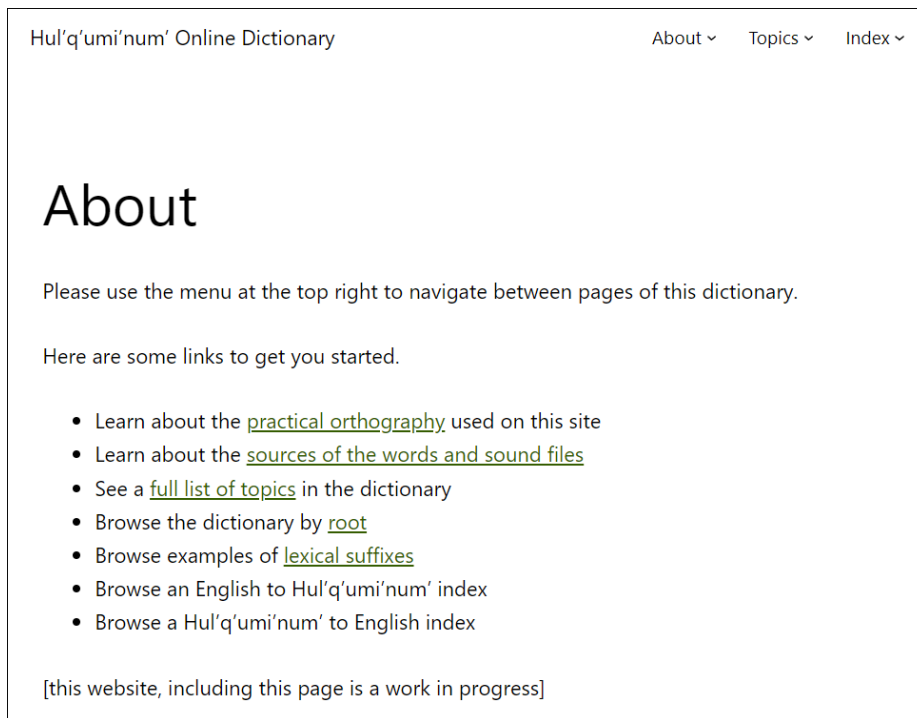


Figure 2.1: Home page

The dictionary website can be found at <https://words.hwulmuhwqun.ca/>. The user navigates the website via the menu at the top right corner. The pages of the website are organized into three groupings - *About*, *Topics*, and *Index*. The remainder of this chapter will detail the individual components of the website followed by an explanation of how they fit into the lexicographic structure of a dictionary.

2.1 Components of the website

In the following section, I describe the main components of the website that the user interacts with. These are the *Topics*, *File card entries*, *Audio*, *Roots*, *Lexical suffixes*, *Audio*, and *Images*. They can be accessed by the menu.

2.1.1 Topics

Under the *Topics* heading on the menu, there are seven major topics (six “worlds” and one section on language basics).



Figure 2.2: Website - Topics

Each of these topic pages are shown in a way so that teachers and learners of Hul’q’umi’num’ can easily find words and hear their sounds. These topics are essentially word lists sorted into categories. Each word on a topic page has a featured audio file that can be played with the button next to the definition.

During the summer of 2019, several meetings with language teachers and students from the HLA took place to decide on a new set of topics for what at the time was

intended to be a childrens' dictionary.⁹ Because the Western-scientific ordering of knowledge tends to fragment Indigenous knowledge (van der Velden, 2010, van der Velden, 2013), the new classification system should be organized by Indigenous concepts and incorporate the language (Cherry and Mukunda, 2015; Pettersen, 2011). As teachers or parents themselves, the members of the HLA could draw on their own experiences to identify the kind of vocabulary that is needed to pass the language down to the younger generation. Collectively they decided to group words around the idea of “worlds” from the perspective of a child. There are six of these worlds: *The World Around Us*, *Us Humans*, *Life at Home*, *Out & About*, *People at Work & Play*, and *Time & Number*. Within each world there are topics and subtopics which act as finding aids. In addition, there is a section called *Language Basics*, which encompasses some of the basic words that transcend categorization, such as function words (pronouns, determiners, etc), basic descriptive words (colors, adjectives), locational words, and the most common verbs. These are the ubiquitous words that form many sentences, and so this section is intended to be a place for people to start when they are first learning the language.

Some words are duplicated in different topics, but since this dictionary is on a website, there is little need to conserve space, unlike with a print dictionary. For example, in the subtopic *Action Words* under *Language Basics*, we have words like *t'ilum* “sing” and *qw'uyulush* “dance”, but we also have those in the category under *People at Work and Play*. The word *t'iwiluh* “pray” is in the *Basic Verbs* but also in the section for *Religion*; *'ulhtun* “eat” is in the *Basic Verbs* as well as *At the Table* (within *Life at Home*) section.

2.1.2 File card entries

While on a Topic page, when the user clicks on a word, a detailed entry opens up in the style of a “file card”. Prior to the widespread use of computers, linguists used file cards or index cards to organize their language documentation. The lexicographic structure of the file card page is described in more detail in Section 2.2.2.

⁹We want to thank Rae Anne Claxton, Martina Joe, Kathleen Johnnie, Donna Modeste, and Sharon Seymour for their work in deciding on the topics and helping to organize words into them.

2.1.3 Roots

A list of roots can be accessed on the website,¹⁰ as shown in Figure 2.3. Clicking on any root will bring the user to a list of file card entries of words associated with that root, with an example shown in Figure 2.4.

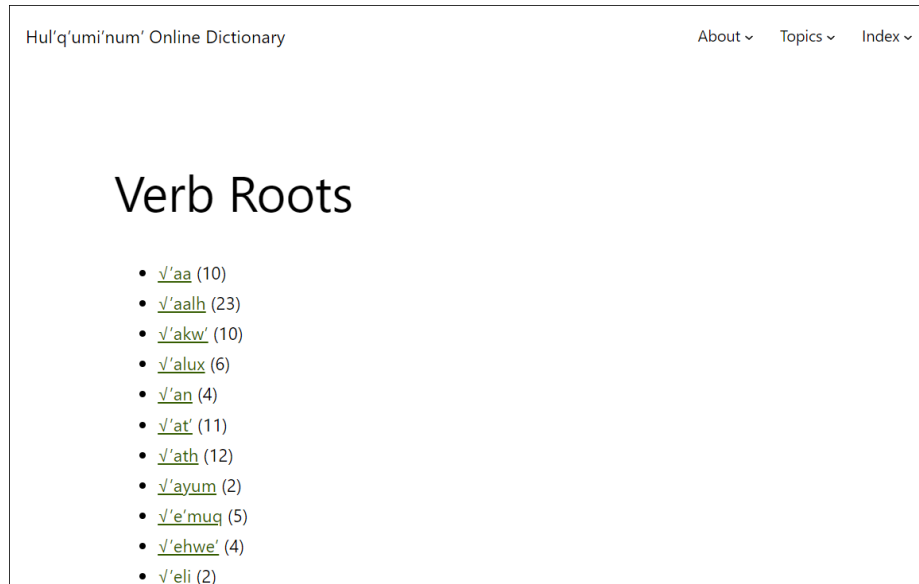


Figure 2.3: Website - Verb roots

The verb system of Hul'q'umi'num' is very different from the verb system of English. There are approximately 680 verb roots in Hul'q'umi'num' [Donna Gerdt, p.c.], and then various derivational and inflectional suffixes are combined with these roots as building blocks to create a variety of meanings. It is helpful for a learner to see related words that are built from the same root grouped together. This also allows them to see how the form of the root changes based on phonological and morphological processes that affect them. Many dictionaries of Salish languages contain a root index (cf. Galloway (2009) and Montler (2018)).

For example, the verb root \sqrt{lem} , which does not occur as a free-standing word without some modification, can have the following meanings in English:

- $ts = lem$ with the verbalizing prefix $ts =$ means “take a look”

¹⁰List of verb roots - <https://words.hwulmuhwqun.ca/index/roots/>

- *lemut* with the transitive suffix = *ut* means “look at it”
- *lumnuhw* with the limited control suffix = *nuhw* means “see it”

Hul'q'umi'num' Online Dictionary About ▾ Topics ▾ Index ▾

Category: √sq'e

<p>yu saq'thut</p> <p>APA: yə saq'θət Parse: yu √sq'e=that Definition: breaking (daybreak)</p> <p>March 23, 2023</p>	<p>suq'suq'utum</p> <p>APA: səqsəqətəm Parse: √sq'e=t=m=PL Definition: split; tear PL</p> <p>March 23, 2023</p>	<p>suq'ul'tsup</p> <p>APA: səqəlçəp Parse: √sq'e=ul'tsup=PROG Definition: splitting wood</p> <p>March 23, 2023</p>
<p>suq'ul'tsupmin't</p> <p>APA: səqəlçəpmin't Parse: √sq'e=ul'tsup=min=t=PROG Definition: splitting, chopping his wood for him</p>	<p>suq'ul'wutum'</p> <p>APA: səqəlwatəm Parse: √sq'e=ulwut=m=PROG Definition: tearing one's clothes</p>	<p>suq'um'utth'ul's stuhw</p> <p>APA: səqəmət'əlsstəx* Parse: √sq'e=emutth'=els=stuhw=PROG</p>

Figure 2.4: Website - Verb root example

2.1.4 Lexical suffixes

Lexical suffixes are suffixes which can attach to words and have properties of nouns (Gerdt, 2003; Gerdt and Hinkson, 2004; Gerdt and Hukari, n.d.). Similarly to roots, many words are derived from the addition of lexical suffixes. By clicking on a lexical suffix, a user can see multiple words with the same lexical suffix.¹¹ They can quickly see how the same lexical suffix changes shape when it combines with different roots and how they contribute to a range of meanings. For example, the lexical suffix = *ew't-hw* adds the meaning of *'building'* when it is added to words (see Table 2.1).

¹¹Lexical suffixes on the website can be browsed on this page - <https://words.hwulmuhwqun.ca/index/lexical-suffixes/>.

Hul'q'umi'num' Online Dictionary About ▾ Topics ▾ Index ▾

Lexical suffixes

=al=qen (4)	=al=we'lh (side) (1)	=als (round object) (34)	
=alus (eye) (29)	=as (face) (116)	=athun (13)	
=ay=thun (=ay=thin) (60)	=ayi (1)	=ay'=a'th (prong) (1)	
=a'luw'=tsus (finger) (7)	=a'qw (head) (98)	=a'th (edge) (14)	
=een (end) (11)	=elu (people) (15)	=elu (place) (1)	=eluqun (1)
=eluqup (3)	=el'ts' (hair) (5)	=emutth' (42)	=enhw (4)

Figure 2.5: Website - Lexical suffixes

Hul'q'umi'num' Online Dictionary About ▾ Topics ▾ Index ▾

Tag: =alus (eye)

<p>yum'alustuhw</p> <p>APA: yəmələstəx* Parse: √yem'=alus=stuhw Definition: wrap up, lace to cradle through eyes on cradle</p> <p style="text-align: right;">March 23, 2023</p>	<p>wuw'tth'a'lust</p> <p>APA: wəw'tʰələst Parse: √wetth'=alus=t=DUR Definition: knitting it (keep on)</p> <p style="text-align: right;">March 23, 2023</p>	<p>wutth'alust</p> <p>APA: wətʰələst Parse: √wetth'=alus=t Definition: knit it</p> <p style="text-align: right;">March 23, 2023</p>
<p>tskwimulus</p> <p>APA: ck*imələs Parse: ts=√kwim=alus Definition: reddish brown</p>	<p>tskwalus</p> <p>APA: ck*aləs Parse: √tsakw=alus Definition: very low tide, water is way out</p>	<p>ts'hwalust</p> <p>APA: çx*aləst Parse: √ts'hwa=alus=t Definition: more: add more (stitches)</p>

Figure 2.6: Website - Lexical suffix example

Table 2.1: Words with =ew't-hw

Word	with =ew't-hw
telu	telew't-hw
<i>money</i>	<i>bank</i>
syalh	yalhew't-hw
<i>firewood</i>	<i>woodshed</i>
q'aq'i'	q'aq'i'ew't-hw
<i>to be sick</i>	<i>hospital</i>

2.1.5 Audio

On the main topic pages, learners can play a featured audio file for each word. When they click on a link that brings them to the file card entry, they can see additional audio files of the same word if available. The audio featured on the website comes from the 1999 Talking Dictionary (Gerds and Penrowley, 1999), language learning materials from the *ta'ulthun sqwal* website,¹² and 2019 audio recordings of Elder Delores Louie. The file card will display the name of the speaker underneath the audio file if that information was entered into the database, otherwise the user will be directed via hyperlink to the full list of elders.

Having audio is extremely important for the dictionary because learners need to hear the language being spoken. The population of Elders who speak the language is small and aging, and at the same time, Hul'q'umi'num' contains many sounds which are difficult for new learners to pronounce.

2.1.6 Images

Some words, especially nouns, may have an image associated with them. At the time of this thesis, the website features photographs by Cim MacDonald, a nature photographer and frequent collaborator with the HLCS, based on Vancouver Island.

¹²ta'ulthun sqwal - <https://sqwal.hwulmuhwqun.ca/>

2.2 Dictionary structure

This section will explain how the components of the website fit into the lexicographic structure of a dictionary (Svensén, 1993, 2009).

2.2.1 Front matter - about the dictionary

At the time of this thesis, the front matter of the dictionary website consists of a page about the practical orthography (see Appendix A) and a page briefly outlining the history of the website and source of materials.

2.2.2 Microstructure - the file card entry

Each “file card” on the website represents a dictionary entry. The title of the file card is the word written in the practical orthography. Inside the file card is some basic information about the word including:

- **the APA phonetic pronunciation**¹³
- **the parse**, which gives the morphological breakdown of a word
- **the definition**

Some additional information can appear on a file card entry depending on the word.

This includes:

- **ethno-biological information**

This may appear on plant and animal entries.

- **a text note**

This may contain additional information about etymology, a usage note, a grammatical note, or a cross-reference.

- **example sentences**

- **an image**

¹³I created a Word macro that transliterated words in the Hul’q’umi’num’ practical orthography to the APA.

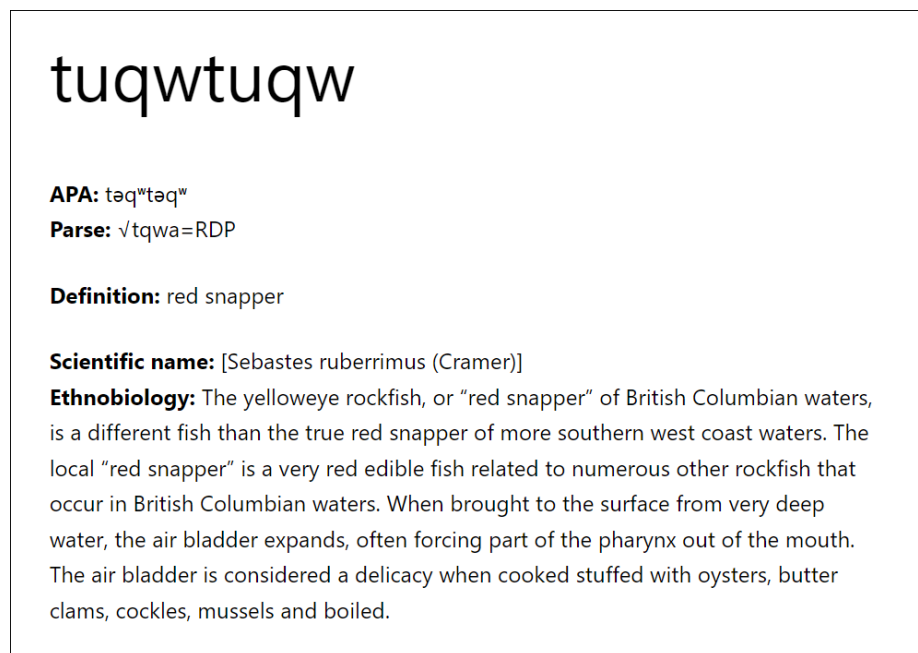


Figure 2.7: Detailed entry

2.2.3 Macrostructure - arrangement of words

As noted above, one of the biggest advantages of electronic dictionaries is that they are not subject to the same space constraints as printed books. This makes it possible to present both systematic (e.g., by topic) and alphabetical arrangements of the same words. In the case of this dictionary, the systematic macrostructure arranges words by semantic topic and sub-topic, and the alphabetical macrostructure is found on the index pages.¹⁴

One of the benefits of arranging words by topic is that nouns, verbs, and descriptives are put together. For example, if the user looks at the topic of *Life at home*, they will find the names of food, the tools or household items associated with food, and verbs relating to preparing or cooking food. Categorization of this kind makes it easier to collect the vocabulary needed to create a lesson plan around specific topics.

2.2.4 Cross-reference structure - linking words

A cross-reference may appear as a text note on the file card view. Cross-reference links are especially important for a learners’ dictionary (Fuertes-Olivera, 2017). Some words

¹⁴The alphabetical order only applies to the first sound. See Figure 3.9.

Table 2.2: Some food related words

Hul'q'umi'num'	Definition
tth'utth'uxals	eggs
stth'ulaalh	jerky (fish or meat)
t'ulum'	cherry: wild cherry
lupat	cup
xe'luw'	spoon (horn or wood), ladle, serving spoon
pi'kwun	roasting stick (split)
t'eets'	cross-sticks (used for smoking salmon, barbecuing meat)
lhts'alsels	cut up apples, vegetables
q'il'aam'	preserve food, make sq'i'lu (drying, canning, etc)
hwlhuw'ut	shuck clams

are commonly confused by learners, and a cross-reference can help learners find the correct word they are looking for. This is especially helpful when there are differences in concept formation between Hul'q'umi'num' and English. In Hul'q'umi'num', the following four verb roots have meanings associated with the concept of breaking something:

- √**t'qw'a** break as in break a piece off something, break or cut something in two
- √**yakw'** break as in smash into tiny pieces
- √**pqwa** break
- √**lukwa** break a stick

Another example is the English word “gather”. In Hul'q'umi'num', one would use *'aluxut* in the sense of gathering plants, shellfish, etc., but *q'apthut* in the sense of people gathering together, e.g. in a meeting.

Cross-reference links can also point the user to a word that shares the same meaning but is more appropriate. For example, the file cards for the words *ha'put*, an older and less commonly used word for deer, contain a cross reference link pointing the user to *smuyuth*, the more commonly used word for deer.

2.2.5 Access structure

The website is intended to be useful for users who do not have formal training in linguistics. It is common for print dictionaries to use different typographical structure markings, such as using different fonts or font sizes, to mark different types of information on an entry. There is enough space on a website to print the full label for terms such as *APA* and *Parse*, so we can print the full label for a field and avoid the use of different typographical structure markings as a distinguishing feature.

2.2.6 Other

The current search features are basic, as sophisticated search features are difficult to implement and so are beyond the scope of the current state of this project. The most effective search procedure is to visit the *Hul'q'umi'num' to English* and *English to Hul'q'umi'num'* index pages which, thanks again to the electronic medium, can be extremely long, where users can use their web browser's built-in search function (i.e., by using the CTRL + F keyboard shortcut) and search the entire list there. There is also a built-in WordPress search bar.

Chapter 3

Building the dictionary

I began my research the old-fashioned way, by visiting libraries and reviewing dictionaries of Salish languages as well as other Indigenous languages. I made photocopies of the front matter and a few pages of entries for future reference. The dictionaries I found useful include Chejn (2016), van Eijk (2013), Galloway (2009), Helmbrecht and Lehmann (2010), Jacobs and Jacobs (2011), Montler (2018), Montler (2012), Pete (2011), and Thompson and Thompson (1996). I also looked at bilingual dictionaries designed for children, available for widely spoken languages such as Spanish, French, or Mandarin Chinese.

The next step was looking into the actual process of compiling a dictionary and producing a version displayed on a website. The idea of making a dictionary website in our situation was complicated by the fact that this project would not be a direct digitization of a singular, pre-existing, finished resource. The materials would have to be organized and compiled from multiple sources in tandem with the creation of a web version. I also needed to establish a workflow connecting the compiled database and the website so the data displayed on the website could be continually updated with any new words. I found that the design of this workflow was just as important as the design of the database or website itself, and I spent as much time researching and designing this component of the project as I did designing the database and the website.

There are many different technical and procedural approaches to dictionary projects, and many examples of Indigenous language dictionaries, online dictionaries, and

community-based dictionaries to look at for guidance and inspiration. I found that the dictionary websites I reviewed fell into three general categories:

- **The dictionary website is created by a database file export.**

Software produced by SIL International, such as Toolbox or Fieldworks, can export *.html files ready for the web (Coward and Grimes, 2000).

- **The dictionary website is completely separate from the lexical database.**

These websites allow words to be entered manually or provide a template for *.csv import. Examples are the FirstVoices¹⁵ website and the Mother Tongues¹⁶ web dictionaries.

- **The dictionary website is fully integrated with a web-based database.**

These are websites which feature both a dictionary for the public and a sophisticated web-based interface for a database for the dictionary team. The dictionary team can directly log-in to the database using the website and make updates there. These changes will then be reflected on the website, pending internal approval. Examples include the Blackfoot Language Atlas (Junker, 2018)¹⁷ and the Swarthmore Talking Dictionaries¹⁸ project.

Another phenomenon I noticed while surveying examples of dictionary websites was that many of them had become outdated, perhaps because they were made at a time when technology standards were less stable. In recent years several standards and formats that had been widely adopted have become obsolete. For example, in the case of Adobe (previously Macromedia) Flash, a platform that enabled certain kinds of interactive web applications, the format became obsolete to the point where modern browsers do not enable backwards compatibility for security reasons. It was extremely

¹⁵FirstVoices - <https://www.firstvoices.com/>

¹⁶Mother Tongues - <https://mothertongues.org/>

¹⁷Blackfoot Language Atlas - <https://blackfoot.algonquianlanguages.ca/>

¹⁸Swarthmore Talking Dictionaries - <http://talkingdictionaries.swarthmore.edu/>

important in designing the dictionary to try to make sure that the technology used would serve the community's needs while being extensible for the future.

Making a plan for the short term was relatively straightforward while I considered the above factors. I decided to continue using a spreadsheet for the time being, because the text data was either already in spreadsheet form or needed to be processed into a spreadsheet. The next task was to get a subset of the words onto a “learners’ dictionary” website on WordPress. WordPress was an easy choice because the HLCS was already using WordPress for all of their websites. At the time, I believed that this website would be a relatively quick project, like a “stepping stone”, and afterwards I would focus on organizing new and legacy data for eventual incorporation into a proper relational database. The initial approach is summarized as such:¹⁹

1. Continue using a spreadsheet in the short-term.
2. Structure the data and organize a subset into a basic learners’ dictionary.
3. Create a website on WordPress and display the dictionary data there.
4. Continue organizing the data to be incorporated into a proper (relational) database for the long-term.

In Section 3.1 I will give an overview of the data when I started working. Then in Sections 3.2, 3.3, and 3.4 I will explain the steps that were taken to organize and prepare the data for the website and database.

3.1 The initial data situation

Thanks to past and ongoing work, I have a wealth of material available to work with, and modern tools to help me organize them. The materials existed in this form when I started:

- **1995 data**

¹⁹While this seemed like a logical sequence of tasks at the time, in reality all the tasks were overlapping and it seemed like all the steps took place simultaneously. This chapter attempts to lay things out in a logical order. See the for a more chronological timeline.

The data which made up the basis for the Hukari and Peter (1995) dictionary was available as a Microsoft Word (MS Word) document (i.e. a .docx file). Past collaborators include Bill Poser, Peter Baer, and Mara Katz who worked to convert this document into a spreadsheet, and Essa Gierc and Janet Leonard also contributed to work on this data. Donna Gerdts, Sally Hart, and Elder Delores Louie added a categorization of *A*, *B*, and *C* words. Words labeled with *A* are appropriate for beginner learners, words with *B* are appropriate for intermediate learners, and *C* represents words for advanced learners. Words that should be considered private were also included in *C*.

- **Hul'q'umi'num' Words Dictionary** (Gerdts et al., 1997)

This was available to me as a MS Word document typed to display APA characters using a legacy font.²⁰ Zoey Peterson, Bill Poser, and Peter Baer had also prepared a transliterated version in the Nanaimo practical orthography.

- **Halkomelem Talking Dictionary** (Gerdts and Penrowley, 1999)

This consisted of approximately 3,500 audio files in *.wav format. Celinda (Cj) Rice did the noise reduction on these files.

- **ta'ulthun sqwal**

The *ta'ulthun sqwal* website²¹ contains numerous language learning materials, including word lists and audio files. This material was already published as a WordPress website. While some of the material existed in the form of MS Word documents, updates and corrections are made directly to the website on a continuing basis in order to support ongoing language classes.

- **2019 recordings of Elder Delores Louie**

²⁰In the earlier days of computing, it was necessary to create fonts to type any characters that were not found in the standard character encodings at the time. The linguist Charles Ulrich created a font, *Straight*, so that people could type APA characters for Salish languages. For more information see Appendix E.

²¹<http://sqwal.hwulmuhwqun.ca/>

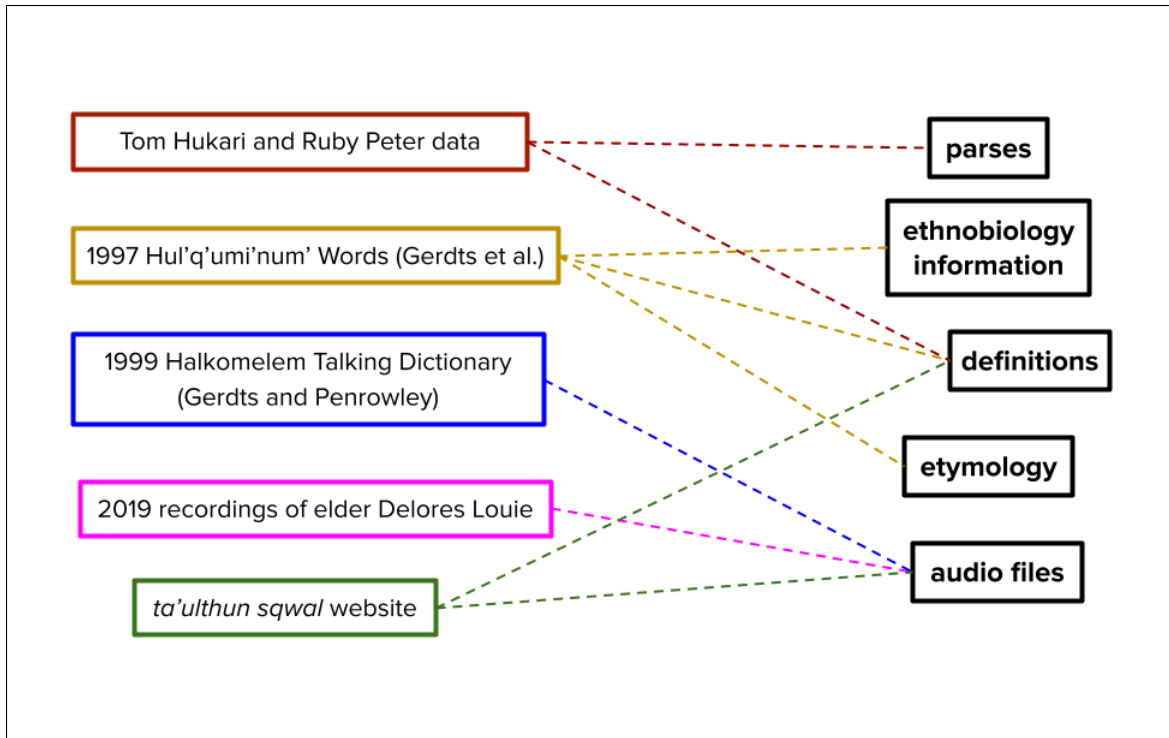


Figure 3.1: Incorporating new and legacy work

During the summer of 2019, approximately 2,000 audio files were recorded of Elder Delores Louie. Thomas Johnny, Evangeline Paige, and Chris Alphonse recorded and segmented the audio files.

- **Photographs**

The nature and wildlife photographer Cim MacDonald provided high quality photographs for us to use in the dictionary.

While I had considerable amounts of material for me to turn into a dictionary website, the idiosyncrasies of the wide range of formats from many different time periods presented many challenges to processing the text for inclusion in a database, as described in the next section.

3.2 Text processing

The legacy documents presented several issues with encoding. A set of MS Word macros known as the *Autoglosser* had been developed by Zoey Peterson which would convert

Word documents typed in APA phonetics (with *Straight* font encoding) into the new practical orthography (with Unicode encoding).²² The Autoglosser also performed basic glossing and formatting, using a small built-in dictionary of approximately 200 commonly glossed words. The transliteration process had two main issues. First, the previously-established process of converting from practical orthography to phonetics yielded a mix of modern Unicode encoding and non-standard encoding, and so I had to update the procedure to address this issue. Second, there were issues with the transliteration of glottalized resonants.

My first major task was to completely revamp the existing transliteration process.²³ I first separated the transcoding task (adherence to the Unicode encoding standard) from the transliteration task (converting between orthographies) by creating a separate set of character mappings for each task. For the transcoding task, there is one MS Word macro that converts documents typed in *Straight* font into the standard Unicode version of the corresponding characters, and for the transliteration task, there are two separate macros that perform the conversions between APA phonetics and practical orthography. The macros that perform the transliteration between APA phonetic characters and the practical orthography require that the target text is in the Unicode standard first. This means that, from now on, any legacy materials typed in *Straight* font must first be transcoded into Unicode, which also ensures that these documents are updated to modern standards.

After several rounds of testing and revising, the transliterator works well. Several students and teachers have been trained on how to use it. A summary of the improvements I made are below.

1. I separated the transcoding task (converting to Unicode encoding standard) from the transliteration task (converting text between different orthographies).

²²I want to thank Zack Gilkison for helping me to understand the Autoglosser and font situation.

²³See Appendices D and E for more details on the Transliteration and Transcoding processes.

2. I added more characters to the *Straight* font transcoder in order to fully transliterate some legacy documents (see Appendix E).
3. I fixed a group of rules pertaining to intervocalic glottalized resonants.

The following list below summarizes the main tasks of the initial text processing of the data from the Hukari and Peter (1995) provided to the HLCS by Tom Hukari. It is a mixture of data cleanup and some minor changes related to the new orthography.

1. **Change outdated translations and spellings**

In rare instances I changed some outdated words and names in the English translations of some words and sentences. Some bands have adopted new spellings of their names to reflect orthography changes or language reclamation efforts. In the case of outdated names I added a note with the former name. For example, I updated the entry for Kuper Island to reflect its current name as Penelakut Island and added the note “*formerly Kuper Island*”.

2. **Update and verify the representation of intervocalic glottalized resonants**

When transliterating from phonetics to practical orthography, a glottalized resonant changes from a “stacked” combination of letter and combining diacritic into two separate characters. For example, $\overset{\text{h}}{\text{m}}$ (stacked) \rightarrow m’ (unstacked). The glottal stop, represented by an apostrophe in the practical orthography, can move to the left or right of intervocalic glottalized resonants (VRV), depending on vowel strength (Gerdt and Hukari, n.d.). Previous transliteration efforts resulted in the glottal stop sometimes being placed on the incorrect side of the resonant. Over the years, research has clarified the behavior of these intervocalic glottalized resonants, allowing me to update the transliteration rules relating to the placement of glottalization for future transliterations. Existing data was corrected manually by using regular expression searches for VRV combinations.

3. Update parse to use the equals sign character ‘=’ instead of the hyphen character ‘-’

The practical orthography uses hyphens to indicate that 2 neighboring letters are not digraphs. In order to avoid confusion between the glossing hyphen and the orthographic hyphen, the parse has been updated to use the equal sign ‘=’ in place of hyphens. Only orthographic hyphens remain in the parse.

4. Fix other minor errors

Minor typos relating to orthography changes or formatting issues.

The Gerdts et al., 1997 dictionary was available to me as a MS Word document. I processed and formatted this into a format that could be imported into a spreadsheet largely by using regular expressions with the *find and replace* function in MS Word.

For the materials on the *ta’ulthun sqwal* website, I started compiling a spreadsheet of words and their associated audio files from the website. The current procedure is to log into the WordPress editor, exporting the code as it exists off of individual pages into MS Excel, and then regularizing it into rows.²⁴

3.3 From spreadsheet to database

Spreadsheet software such as MS Excel and Google Sheets was useful in cleaning up and organizing the legacy data, but became difficult to manage when it came time to incorporate new words and sound files. Although intuitive to many users, spreadsheets are ultimately limited in power and structure (Wattanawaroon, 2021). Most importantly, it was crucial to have a relational database that could properly handle complicated relationships in the data that are not one-to-one, such as many words being linked to one example sentence (a many-to-one relation), or one word having multiple definitions (a one-to-many relation).

²⁴At the time of this thesis I have processed the pages in the *kwunthat ch - Getting started* section and the *’ilhe hwulmuhwqun – Language lessons* sections of the website.

There are several software solutions for lexicography on the market, including TshwaneLex,²⁵ Miromaa,²⁶ SIL Fieldworks,²⁷ and more. Each solution would involve a significant investment of time and effort to make our data and our situation fit into a mold.²⁸ In fact, the “standardised data approach” represented by one-size-fits-all solutions may be “preventing 21st century lexicography from making full use of the new computer and information technologies” (Tarp, 2015, p. 219).

In order to reconcile the need for an intuitive interface with the need for the complexity and power of relational databases, I decided on Airtable, a “hybrid spreadsheet” cloud platform that has features of both spreadsheets and relational database management systems (RDBMS). It is an example of a low-code development platform (LCDP)²⁹ that uses visual interfaces and menus to enable people with limited technical skills to develop applications. This results in a significantly lower cost of learning compared to traditional database management systems, which can “empower users to engage actively in the continuous development of systems rather than being restricted to the use of existing systems” (Fischer et al., 2004, p. 35).

I have found that the main benefits of using Airtable for our project are:

- **Structured data fields and relations**

There are several data types for *fields*, (i.e. columns), including *Link to another record* and *Lookup*. This enables the database to structure relations between forms and meanings with relational-style queries, something which is not well-suited for MS Excel.³⁰ In addition, Airtable *fields* are structured, e.g. anything entered into a *Text* field will always be interpreted as text. There are several characters used in

²⁵TshwaneLex - <https://tshwanedje.com/tshwanelex/>

²⁶Miromaa - <https://www.miromaa.org.au/>

²⁷SIL Fieldworks - <https://software.sil.org/fieldworks/>

²⁸I even considered the idea of designing my own database with web-interface similar to Dunham (2014), an idea I thankfully did not spend too much time on.

²⁹Low-code or no-code are terms used somewhat interchangeably in the literature.

³⁰It is possible to mimic this functionality in an ad-hoc manner with VLOOKUP formulas and other operations. I do not advise it.

our database which are coded as punctuation or trigger formula operations (such as the equal sign or hyphen), and this sometimes causes undesired behavior in spreadsheet software.

- **Graphical user-interface**

A user-friendly interface is extremely important not only for the people looking at the final end-product, but also for the people designing the dictionary database. Traditionally, domain experts (i.e. people who know the language) have been removed from system designers (i.e. people who know the technology) due to the high cost of learning database design (Arias et al., 2000; Tarp, 2019).

- **Custom views**

Custom database *views* can be designed in Airtable to only show the relevant and useful fields for carrying out a specific task. This contextualizes the information to “help people attend to the information that is the most relevant for their task at hand” (Arias et al., 2000, p. 89) and gives the user confidence to work on the database. A well-designed interface for the database will “facilitate the lexicographer’s job, reduce the number of mistakes, economise on the resource employed, and shorten the total production time” (Tarp, 2015, p. 234). Specific views have been setup for tasks such as tagging lexical suffixes and linking sound files to words.

- **Collaboration and access controls**

As Airtable is a cloud-based service, multiple people can work on the same database simultaneously, and they can use computers with Mac or Windows operating systems. *Access control* allows for different levels of users to have different editing permissions, something which helps limit the potential for mistakes.

- **Dashboards**

Dashboards are interface tools common in many applications used to visualize, analyze, and track data in a single place in real-time. The Airtable interface

designer has made it easy to create self-updating dashboards which allow the team to measure progress and see gaps that need to be filled in.³¹

- **Data portability**

The portability of data is its ability to be changed into a different format, moved out of a particular application, used by other applications, etc.³² The database in Airtable can be easily exported as a *.csv file, a common format that is compatible with the workflow I designed to import dictionary data onto the website.

- **Sustainability**

Finally, given that the HLCS currently has no permanent IT staff, it made sense to use software with existing documentation so that the dictionary project is not dependent on a single persons' expertise and idiosyncratic coding style.

The intention of this thesis is not to advocate for a particular software offered by a particular company, which brings up the concerns of data portability (addressed above) and vendor lock-in. To address the concern of vendor lock-in,³³ I created an “everything” view in Airtable, which pulls together all fields through linked table and look-up fields. This view replicates the look of how the data looked previously on Google Sheets. At any point in time, someone can export this view as a .csv file and import it back into Google Sheets, which was a component of a previous workflow that some users are familiar with. This gives us a sense of security that we can always revert to previous ways of doing things, if necessary. Going forward, we must continue to carefully evaluate the pros and cons of using Airtable-specific features such as the interface designer or certain extensions, as we want to avoid being too dependent on them. For more discussion on this, see Section 4.2.4.

One of the benefits of using Airtable is that it has also been relatively easy to make changes to the database so we can respond to the dynamic needs of users. One successful

³¹See Figure 3.6 for an example.

³²See Bird and Simons (2003) for a detailed explanation of data portability.

³³Vendor lock-in is the idea that you can be essentially forced to continue using a product by a specific vendor because switching away is not practical or easy.

case of this responsiveness was in a classroom context. During the fall semester of 2022, Hul'q'umi'num' students in Donna Gerdt's *LING 431: Word Architecture* class were assigned to research specific verb roots and lexical suffixes. In order to be more useful for this and future classes, a "tagging" functionality was added to the database which enabled students to navigate the online version of the dictionary and find words which contained a particular lexical suffix. This was done by adding a linked field into the main table which created a link to a table of lexical suffixes. When the database user wanted to "tag" a word, they clicked on the appropriate field (see Figure 3.2), and this opened a menu where they were able to choose from the full list of lexical suffixes (see Figure 3.3). Students could filter the database and then manually tag words by adding them from a pre-populated menu.

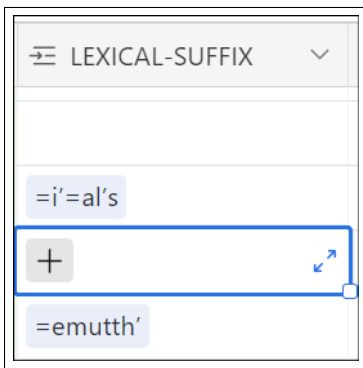


Figure 3.2: Lexical suffix workflow part 1

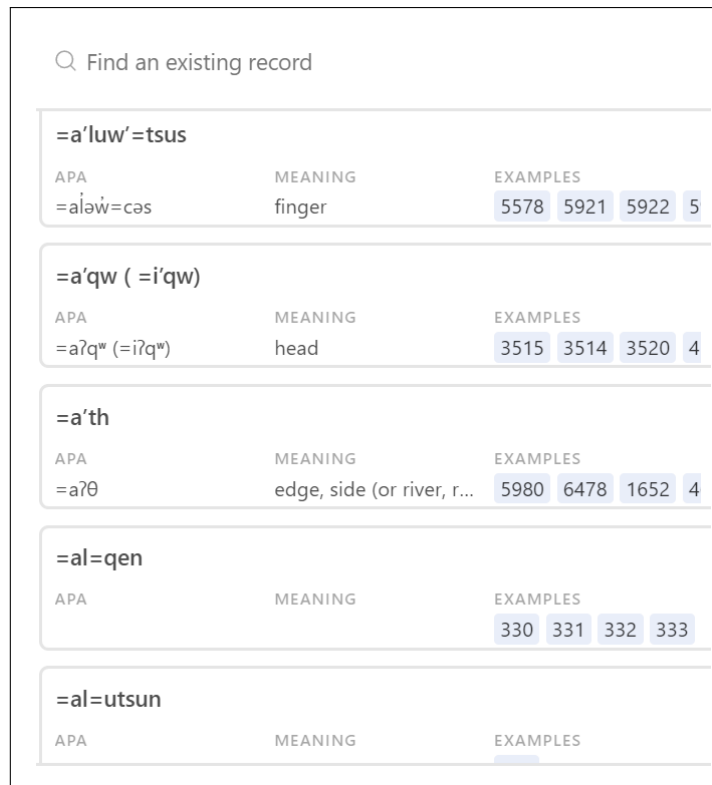


Figure 3.3: Lexical suffix workflow part 2

There are over 70 lexical suffixes in the dictionary, and many more to be added. Of the nearly 6,600 words in the dictionary, approximately 25% have been tagged with a

lexical suffix. This new procedure allows for language workers to continue this important task.

Another feature that has proven to be useful is how Airtable handles audio files. They can be attached to the database and played directly, as shown in Figure 3.4. This makes it easier for the dictionary team to link sound files with words. In many cases, the filename consists of an English word that can act as a ‘hint’ for what the word is in Hul’q’umi’num’. By creating a link between the main lexical table and the audio table, a member of the dictionary team can search for a word and see multiple results before choosing a match. This makes it much easier to sort through audio files of similar words. For example, in Figures 3.4 and 3.5 there are three audio files with the names “BARK1.wav”, “BARK2.wav”, and “BARK3.wav”. The dictionary team member can (1) listen to the audio, (2) search for the word “bark”, and (3) choose the Hul’q’umi’num’ word that matches what they heard without leaving the Airtable interface.

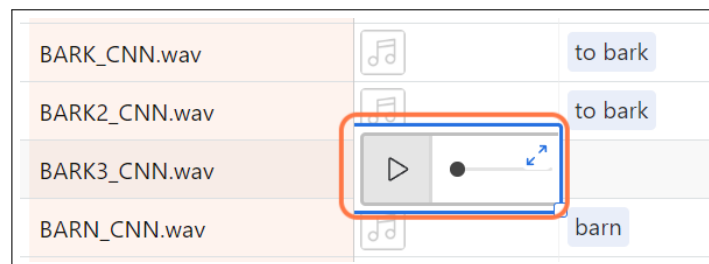


Figure 3.4: Playing audio

As mentioned above, dashboards are a useful tool for the dictionary team to measure progress and see what gaps need to be filled in. They are easy to design and create using the Airtable *Interface designer*. Figure 3.6 shows a dashboard designed to give an overview of how many audio files and images are in different topics of the dictionary. The dashboard can be filtered by topic and subtopic.

3.4 From database to website

The basic elements of a WordPress site include *Posts*, *Pages*, *Categories*, and *Tags*. They are mapped to the dictionary structure in Table 3.1.

🔍 bark | ✕

tree bark	DEFINITION (FROM NO...)	HULQ (LOOKUP)	AUDIOLINK	AUDIO
tree bark		p'uli'		
cedar bark rope	DEFINITION (FROM NO...)	HULQ (LOOKUP)	AUDIOLINK	AUDIO
cedar bark rope		syukw'um		
(seal) to bark	DEFINITION (FROM NO...)	HULQ (LOOKUP)	AUDIOLINK	AUDIO
(seal) to bark		qwelq'um		
inner cedar bark	DEFINITION (FROM NO...)	HULQ (LOOKUP)	AUDIOLINK	AUDIO
inner cedar bark		sluwi'		
wild cherry bark	DEFINITION (FROM NO...)	HULQ (LOOKUP)	AUDIOLINK	AUDIO

+ Add new record

Figure 3.5: Results for “bark”



Figure 3.6: Dashboard

Table 3.1: WordPress to Dictionary mapping

WP Element	Dict Component	Dictionary Function
Page	topical word list	topical dictionary for learners
Post	individual entry	detailed entry with example sentences, image, source information, etc.
Category	Root	root dictionary
Tag	Morphological labels	look-up words by lexical suffix, affixes, etc

3.4.1 Website design

The dictionary website is designed to cater to two different user groups: the *casual learner*, and the *advanced learner or researcher*. WordPress provides a simple *Page* function which works like a common web page creation tool, where the page editor manually chooses and designs what goes on the page. The “learners’ dictionary” aimed at the casual learner currently consists of 7 topic *Pages*, where each *Page* is a “mini dictionary” corresponding to that topic, following the approach of Mosel (2011). These pages are created manually using a Markdown table converter that converts spreadsheet data into Markdown to be displayed on a web page.³⁴ Headings and subheadings were added manually. There are approximately 1,800 words on the learners’ dictionary *Pages*.

Casual learner	Advanced learner or researcher
The casual learner is someone who will navigate by topic .	The advanced learner or researcher is someone who will navigate the dictionary by alphabetical index , or by exploring the verb roots or lexical suffixes .
They will browse words and listen to the featured audio file without necessarily clicking into the detailed “file card”.	They will look at the detailed “file card” entries to find additional audio or example sentences.
1,800 words	7,500 words

Figure 3.7: User groups

³⁴TableConvert - <https://tableconvert.com/excel-to-markdown>

For the advanced learner or researcher, the WordPress *Posts*, *Categories*, and *Tags* are used to organize the dictionary data. WordPress *Posts* (see: “file cards”) are created in a bulk import process using the *WP All Import* plug-in.³⁵ First, the “Everything” view from the Airtable dictionary database is first exported as a .csv file. Then, this .csv file is imported such that each individual record (i.e. a database row) is imported onto the WordPress site as a *Post*. Each *Post* is assigned a *Category* according to the root of the word, and *Tags* are assigned if there are any lexical suffixes are tagged for that word. Users can then find words with a certain root by looking up that root’s *Category* as well as finding a word by looking up its lexical suffix as a *Tag*. There are approximately 7,500 such *Posts*.

3.4.2 Alphabetical ordering

The *Hul’q’umi’num’ to English* index page features words in alphabetical order, based on the Hul’q’umi’num’ practical orthography. The alphabetical order has been organized by the Hul’q’umi’num’ language teachers to resemble that of English.³⁶ The entries in the alphabet represent sounds in the language, and some sounds can be represented by more than one letter. This includes digraphs such as *sh* for /š/, trigraphs such as *kw’* for /kʷ/, and quadgraphs such as *tth’* for /tʰ/. When these sounds occur, they are treated as a single unit for the purpose of alphabetization. The glottal stop is treated as the first letter of the alphabet. After the first sound, the alphabetical order follows the English alphabetical order according to each individual character and not the sound.

³⁵WP All Import - <https://www.wpallimport.com/>

³⁶Hear the alphabet song on the *ta’ulthun sqwal* site - <https://sqwal.hwulmuhwqun.ca/learn/learning-the-sounds/alphabet-song/>.

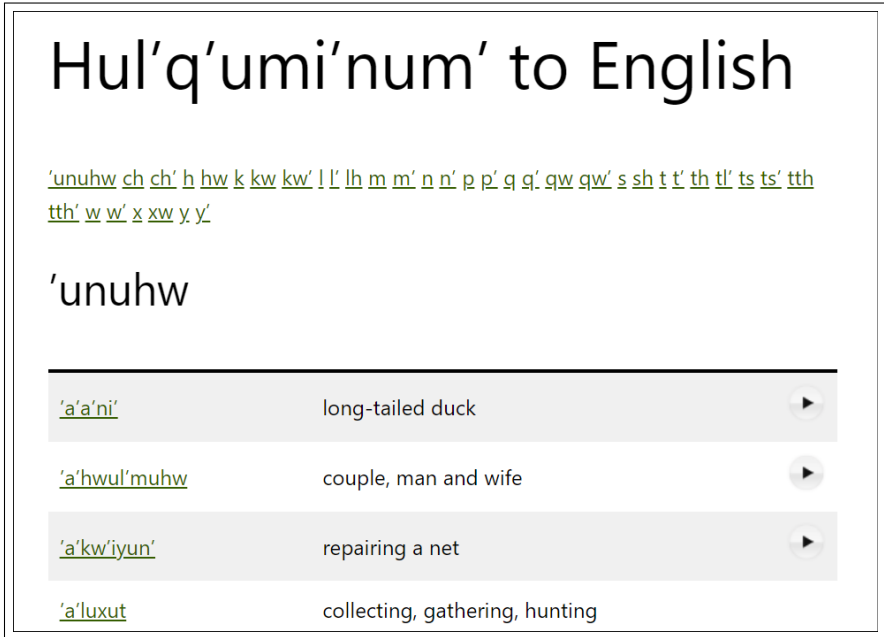


Figure 3.8: Website - Index

```

if the left 3 characters have an exact match in the alpha column; then
  | assign alpha order;
else
  | if the left 2 characters have an exact match in alpha column; then
  | | assign alpha order;
  | else
  | | if the left 1 character has an exact match in alpha column; then
  | | | assign alpha order;
  | | end
  | end
end

```

Figure 3.9: Alphabetical ordering of the first sound

Chapter 4

Conclusion

Now that I have described the interface of the dictionary website and the process of its creation, I conclude here with some personal observations about the non-technical aspects of community-based dictionary projects.

4.1 Reflecting back

Over the course of the past four years of this project, I have learned a lot about working on a dictionary project. At times I was unprepared for some of the realities of working collaboratively, and of trying to manage such a project.

4.1.1 A note on collaborative and participatory models of research

All HLCS projects emphasize being collaborative and participatory. This is the foundation of their collective effort in strengthening the Hul'q'umi'num' language. Collaborative and participatory research can be rewarding and beneficial (Fitzgerald, 2021; Junker, 2018; Stenzel, 2014). Bringing in people to see the process helps everyone gain a better perspective about where their work is going, and “informed participation leads to ownership and a stronger sense of community” (Arias et al., 2000, p. 90).

That being said, in practice it takes a lot of thought and intention to create sustainable participatory roles (Fitzgerald, 2018, 2021; Junker, 2018). Muller and Druin (2012) note that in order to “bring users’ knowledges and perspectives directly into computer specification and design, it is necessary to do more than ‘just add users and stir’” (p. 1125). Introducing any and every person into this type of process without

considering their individual interests can result in mutual frustration instead of mutual learning.

For these reasons, it is important to approach the design of the dictionary as a design problem. If we value participation, we have to explicitly design for it (Fischer et al., 2004). Design should be a continuous process of action and reflection, where people act until they experience a breakdown, and then they reflect (Arias et al., 2000). This reflection is where actual problem-solving occurs. It takes a lot of time to understand how to effect meaningful change, and it is important to be flexible and let priorities change as the design evolves.

4.1.2 A note on project management

Working on any dictionary project involves dealing with “problems in a variety of fields that are not directly connected with the content of the dictionary” (Svensén, 2009), problems which I will group together under the umbrella of “project management”. I often found myself unprepared for these types of problems, and here I will give some advice for anyone in the same situation.

In addition to the usual network contacts that arise organically in this work, I would advise others to try to consult with a few people outside of linguistics, language revitalization, or academia for technical and operational guidance. On one hand, I found that people outside of these spaces have a limited understanding of some of the challenges and limitations that I had to work with, and may provide suggestions that are unfeasible. However, they can also bring a fresh perspective on things.

It is extremely important to make sure that everyone involved in a dictionary project is understanding each other. Technical language can be a huge barrier in communication. Sometimes a database is a relational database, and sometimes a database is a spreadsheet. Do not be afraid to ask for clarification at any time.

Make sure that any instructions are well-defined before diving into a task and that proper support is provided. Spending a little extra time on the planning and design process early on can save time later. Related to this point is the issue of scope creep. Just

because something sounds easy to implement doesn't mean that it is. Make sure that any deviations from the original plan are worth the extra effort and time.

4.2 Looking ahead

Next steps for the future of the dictionary project will be discussed here.

4.2.1 More dictionary content

One goal is to integrate more explanations onto the dictionary website. The people using such a dictionary may not have a formal background in language structure, and it is important to provide explanations about the terminology and concepts that are printed in an entry (Gawne et al., 2017; Kell, 2014). For example, the parse can be a source of confusion for people that have never seen such a thing.

There are many words that need to be added to the database, including many words from the *ta'ulthun sqwal* website, as well as words that appear in stories,³⁷ MA projects, etc. Adding these words is a priority because a lot of them are words commonly encountered by new readers. Another way to gather more words would be to provide a fillable form designed to collect submissions from people which can then be checked by the dictionary team.

4.2.2 Capacity building

More than a dozen people helped with this version of the dictionary project. In the future, others will join the project. It is important for the dictionary database and website to be sustainable. This can be done by having participatory roles and training members of the community (Fitzgerald, 2018). It is also a goal to train students on general spreadsheet and data management skills, which can transfer to the workplace. Students can take on tasks such as adding images to the database or tagging lexical suffixes. Members of the team have many interesting ideas about incorporating cultural information into the

³⁷See Alphonse et al. (2021) for an example.

dictionary. Ultimately, the goal is to encourage more students to become designers and codevelopers, not just consumers (Arias et al., 2000; Fischer, 1998).

4.2.3 Re-use and extensibility

The database work should be seen as a continuous work in progress, and this includes the design. The database “cannot be completely designed prior to use” and “must evolve at the hands of the users” (Arias et al., 2000). Fields can be added for additional types of information depending on future projects.

4.2.4 Future data migration

There is a long history in language documentation of formats, applications, technologies, and entire ways of thinking becoming obsolete (Stevens et al., 2010; Walter and Suina, 2019). It is important to keep in mind the future of the data that we put in any system and here I provide five examples for consideration.

1. **Airtable**, as discussed above.
2. **Open source alternatives**

NocoDB³⁸ and Baserow³⁹ connect to a database and an Airtable-like interface for the end user. At the time of this thesis, both applications were too new for me to consider using in production. Dataspread⁴⁰ has an interface that more closely resembles traditional spreadsheet software.

³⁸NocoDB - <https://www.nocodb.com/> is at v0.105.3 at the time of this thesis.

³⁹Baserow -<https://retool.com/> is at v1.15 at the time of this thesis.

⁴⁰Data spread - <http://dataspread.github.io/>

3. Commercial alternatives

Google Tables⁴¹, Rows.com⁴², and Smartsheet⁴³ are commercial applications with some overlapping features of Airtable.

4. An entirely custom solution

In the future, circumstances may change so that the idea of a completely custom dictionary database and interface becomes possible and desirable.

5. Something entirely different

As technology is continuously evolving, we must continuously re-evaluate our data situation. Future innovation in spreadsheets might not necessarily lead us to another “Airtable clone”.

4.3 Final thoughts

The recent FPCC report on the status of B.C. First Nations languages (Gessner et al., 2022) underscores how important it is to support opportunities for learning First Nations languages. Working on HLCS projects has been a constant struggle to balance needs and priorities with limited resources, a problem Mosel (2011) describes as “trying to square the circle”. At any given time, there are numerous credit and non-credit classes, material development projects, documentation and research projects, and events taking place around the HLCS. HLCS projects aim to strengthen language and culture, and these require resources in terms of people, time and money.

It has been a great privilege to be able to work alongside the Hul’q’umi’num’ teachers, learners, and researchers in their mission of strengthening the language. I hope that technology can continue to provide interesting and novel ways to support their efforts to become more fluent in the language. In particular, I hope this dictionary project

⁴¹Google Tables - [https://en.wikipedia.org/wiki/Tables_\(Google\)](https://en.wikipedia.org/wiki/Tables_(Google)) is in beta and not available in Canada at the time of this thesis.

⁴²Rows.com - <https://rows.com/>

⁴³Smartsheet - <https://www.smartsheet.com/>

will be helpful to the new speakers as well as be of use to other linguists who seek to use technology as a tool to assist communities in revitalizing their languages.

Bibliography

- Alphonse, C., George, R., Joe, M., Johnny, T., Modeste, D., Seymour, A. V. S., & Zhang, H. Y. (2021). Analyzing a story by Sti'tum'at: Tth'uwxe'le'ts - Basket Ogress. In D. K. E. Reisinger, H. Green, L. Griffin, M. Huijsmans, G. Mellesmoen, & B. Trotter (Eds.), *Papers for the 56th International Conference on Salish and Neighbouring Languages*.
- Arias, E., Eden, H., Fischer, G., Gorman, A., & Scharff, E. (2000). Transcending the individual human mind—creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction*, 7(1), 84–113. <https://doi.org/10.1145/344949.345015>
- Berez-Kroeker, A. L., McDonnell, B. J., Koller, E., & Collister, L. B. (2021). *The Open Handbook of Linguistic Data Management*. MIT Press. <https://doi.org/10.7551/mitpress/12200.001.0001>
- Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, 79(3), 557–582. <http://www.jstor.org/stable/4489465>
- Cablitz, G. (2011). The making of a multimedia encyclopaedic lexicon for and in endangered speech communities. In G. Haig, N. Nau, S. Schnell, & C. Wegener (Eds.), *Documenting Endangered Languages* (pp. 223–262). De Gruyter. <https://doi.org/10.1515/9783110260021.223>
- Callaghan, C. A. (2002). Writing a user-friendly dictionary. In W. Frawley, K. C. Hill, & P. Munro (Eds.), *Making Dictionaries: Preserving Indigenous Languages of the Americas* (pp. 312–321). University of California Press.
- Charlie, S. (2019). Kinship terms within the Hul'q'umi'num' territory (Master's project). Simon Fraser University.
- Chejn, A. (2016). System for Creation of a Dictionary: Icelandic-Czech Students' Dictionary. *International Journal of Information and Communication Technologies in Education*, 1(2), 80–86. <https://doi.org/10.1515/ijicte-2012-0007>
- Cherry, A., & Mukunda, K. (2015). A Case Study in Indigenous Classification: Revisiting and Reviving the Brian Deer Scheme. *Cataloging & Classification Quarterly*, 53(5-6), 548–567. <https://doi.org/10.1080/01639374.2015.1008717>
- Claxton, R. A. (2020). *lhwet tse' xwi'em'? hwi' 'een'thu tse'. How I learned to perform a Hul'q'umi'num' story* (Master's thesis). <https://theses.lib.sfu.ca/file/thesis/6072>

- Coward, D. F., & Grimes, C. E. (2000). *Making Dictionaries: A guide to lexicography and the Multi-Dictionary Formatter*. SIL International.
<https://software.sil.org/shoebox/mdf/>
- Cowichan Tribes (Ed.). (2007). *Quw'utsun Hul'q'umi'num' Category Dictionary*. Cowichan Tribes.
- Dunham, J. R. W. (2014). *The Online Linguistic Database: Software for Linguistic Fieldwork* (Doctoral dissertation). University of British Columbia.
<https://doi.org/10.14288/1.0165582>
- Dziemianko, A. (2017). Electronic dictionaries. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography*.
- van Eijk, J. (2013). *Lillooet-English Dictionary*. University of British Columbia, Department of Linguistics. https://lingpapers.sites.olt.ubc.ca/files/2018/01/Van_Eijk_Lillooet-English-Dictionary1-1.pdf
- Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A. G., & Mehandjiev, N. (2004). Meta-design: a manifesto for end-user development. *Commun. ACM*, 47(9), 33–37.
<https://doi.org/10.1145/1015864.1015884>
- Fischer, G. (1998). Beyond “couch potatoes”: From consumers to designers and active contributors. *Asian Pacific Computer Human Interaction Conference (APCHI'98)*, 7, 2–9. <https://doi.org/10.5210/fm.v7i12.1010>
- Fitzgerald, C. M. (2018). Creating sustainable models of language documentation and revitalization. In *Insights from Practices in Community-Based Research* (pp. 94–111). De Gruyter Mouton. <https://doi.org/10.1515/9783110527018-005>
- Fitzgerald, C. M. (2021). A framework for language revitalization and documentation. *Language*, 97(1), e1–e11. <https://doi.org/10.1353/lan.2021.0006>
- Frawley, W., Hill, K. C., & Munro, P. (Eds.). (2002). *Making Dictionaries: Preserving Indigenous Languages of the Americas*. University of California Press.
- Fuertes-Olivera, P. A. (Ed.). (2017). *The Routledge Handbook of Lexicography*. Taylor & Francis Group.
- Galloway, B. D. (2009). *Dictionary of Upriver Halkomelem* (Vol. 1). University of California Press. <https://escholarship.org/uc/item/65r158r4>
- Gawne, L., Kelly, B. F., Berez-Kroeker, A. L., & Heston, T. (2017). Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation*, 157–189.
<http://scholarspace.manoa.hawaii.edu/handle/10125/24731>

- Gerdts, D. B. (1988). *Object and Absolutive in Halkomelem Salish*. Routledge.
- Gerdts, D. B. (1996). *ʔiʔte? xʷəlməxʷqən: tə nəcəm at pukʷ (Let's Speak Our First Nations Language: Book One)*. <http://www.sfu.ca/~gerdts/papers/HalkEd.htm>
- Gerdts, D. B. (1997). *ʔiʔte? xʷəlməxʷqən: tə yəsálam at pukʷ (Let's Speak Our First Nations Language: Book Two)*. <http://www.sfu.ca/~gerdts/papers/HalkEd.htm>
- Gerdts, D. B. (2003). The Morphosyntax of Halkomelem Lexical Suffixes. *International Journal of American Linguistics*, 69(4), 345–356. <https://doi.org/10.1086/382736>
- Gerdts, D. B. (2010). Ditransitive constructions in Halkomelem Salish: A direct object/oblique object language. In A. Malchukov, M. Haspelmath, & B. Comrie (Eds.), *Studies in Ditransitive Constructions* (pp. 563–610). De Gruyter Mouton. <https://doi.org/10.1515/9783110220377.563>
- Gerdts, D. B., Edwards, L., Ulrich, C. H., & Compton, B. D. (Eds.). (1997). *Hul'q'umin'um' Words: An English-to-Hul'q'umin'um' and Hul'q'umin'um'-to-English Dictionary*. <http://www.sfu.ca/~gerdts/papers/HalkEd.htm>
- Gerdts, D. B., Grant, E., & Point, D. (1998). *Musqueam Lessons*.
- Gerdts, D. B., & Hinkson, M. Q. (2004). Salish numeral classifiers: a lexical means to a grammatical end. *Language typology and universals*, 57(2-3). <https://doi.org/10.1524/stuf.2004.57.23.247>
- Gerdts, D. B., & Hukari, T. E. (n.d.). *Hul'q'umi'num': A Grammar of Island Halkomelem* [Unpublished manuscript].
- Gerdts, D. B., & Hukari, T. E. (2000). Stacked Antipassives in Halkomelem Salish. *Papers for the 35th International Conference on Salish and Neighboring Languages*, 95–106. https://lingpapers.sites.olt.ubc.ca/files/2018/03/2000_Gerdts_Hukari.pdf
- Gerdts, D. B., & Hukari, T. E. (2004). Determiners and transitivity in Halkomelem texts. In D. B. Gerdts & L. Matthewson (Eds.), *Studies in Salish linguistics in honor of M. Dale Kinkade*. (pp. 150–170). University of Montana Occasional Papers in Linguistics. https://www.sfu.ca/~gerdts/papers/Gerdts_Hukari_KinkadeFest.pdf
- Gerdts, D. B., & Hukari, T. E. (2006). The Halkomelem Middle: A Complex Network of Constructions. *Anthropological Linguistics*, 48(1), 44–81. <http://www.jstor.org/stable/25132365>
- Gerdts, D. B., & Hukari, T. E. (2011). The Dual Structure of Halkomelem Motion Verbs. *Northwest J. Linguist.*, 5, 1–20.
- Gerdts, D. B., & Penrowley, C. (1999). *Halkomelem Talking Dictionary*.

- Gerds, D. B., & Werle, A. (2014). Halkomelem clitic types. *Morphology*, 24(3), 245–281. <https://doi.org/10.1007/s11525-014-9241-0>
- Gessner, S., Herbert, T., & Parker, A. (2022). Report on the status of B.C. First Nations Languages, Fourth Edition 2022 (fourth edition). <https://fpcc.ca/wp-content/uploads/2023/02/FPCC-LanguageReport-23.02.14-FINAL.pdf>
- Hargus, S. (2019). The Witsuwit'en and Kwadacha dictionary projects. https://faculty.washington.edu/sharon/AAA-CASCA_Hargus-UW.pdf
- Helmbrecht, J., & Lehmann, C. (Eds.). (2010). *Hoçak Teaching Materials: Elements of Grammar/Learner's Dictionary* (Vol. 1). State University of New York Press.
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36(1), 161–196. <https://doi.org/10.1515/ling.1998.36.1.161>
- Hinton, L., & Weigel, W. F. (2002). A Dictionary for Whom? Tensions between academic and nonacademic functions of bilingual dictionaries. In W. Frawley, K. C. Hill, & P. Munro (Eds.), *Making Dictionaries: Preserving Indigenous Languages of the Americas* (pp. 155–170). University of California Press.
- Hukari, T. E. (1978). Halkomelem nonsegmental morphology. *Papers for the 13th International Conference on Salish and Neighboring Languages*, 157–207.
- Hukari, T. E. (1981). Some phonological problems in the Cowichan l-infix plural. *Working Papers for the XVI International Conference on Salish Languages*, 70–79.
- Hukari, T. E., & Peter, R. (Eds.). (1995). *The Cowichan Dictionary of the Hul'q'umin'um' Dialect of the Coast Salish People*. Cowichan Tribes.
- Jacobs, P., & Jacobs, D. (Eds.). (2011). *Skwxwú7mesh Sníchim Xwelíten Sníchim: Skexwts*. University of Washington Press. <https://uwapress.uw.edu/book/9780295990224/squamish-english-dictionary/>
- Junker, M.-O. (2018). Participatory action research for Indigenous linguistics in the digital age. In *Insights from Practices in Community-Based Research* (pp. 164–175). De Gruyter Mouton. <https://www.degruyter.com/document/doi/10.1515/9783110527018-009>
- Kell, S. (2014). *Polysynthetic Language Structures and their Role in Pedagogy and Curriculum for BC Indigenous Languages* (tech. rep.). Aboriginal Education Team, BC Ministry of Education. https://www.llbc.leg.bc.ca/public/pubdocs/bcdocs2014_2/545041/polysynthetic_language.pdf
- Leslie, A. R. (1979a). *Dictionary of the Nanaimo Dialect (Coast Salish)*. Compiled from 1970 Field Notes.

- Leslie, A. R. (1979b). *A Grammar of the Cowichan Dialect of Halkomelem Salish* (Doctoral dissertation). University of Victoria.
- Lew, R. (2014). Dictionaries and their users. In P. Hanks & G.-M. de Schryver (Eds.), *International Handbook of Modern Lexis and Lexicography* (pp. 1–9). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-45369-4_11-1
- Luning, R., & Yamauchi, L. (2010). The influences of indigenous heritage language education on students and families in a Hawaiian language immersion program. *Special Issue on Identity*, 7(2), 207–236. <https://doi.org/10.46538/hlj.7.2.4>
- Marston, L. (2021). xu’athun stl’q’een’: hwulmuhw perspectives on art and language (Master’s project). Simon Fraser University.
- Montler, T. (Ed.). (2018). *SENĆOŦEN: A Dictionary of the Saanich Language*. University of Washington Press.
- Montler, T. (Ed.). (2012). *Klallam Dictionary*. University of Washington Press.
- Mosel, U. (2011). Lexicography in endangered language communities. In *The Cambridge Handbook of Endangered Languages* (pp. 337–353). Cambridge University Press. <https://doi.org/10.1017/CBO9780511975981.017>
- Muller, M. J., & Druin, A. (2012). Participatory design: The third space in human–computer interaction. In *The Human–Computer Interaction Handbook* (pp. 1125–1153). CRC Press.
- Nathan, D., & Fang, M. (2013). Re-imagining documentary linguistics as a revitalization-driven practice. In *Keeping Languages Alive: Documentation, Pedagogy and Revitalization* (pp. 42–55). Cambridge University Press. <https://doi.org/10.1017/CBO9781139245890.006>
- Nielsen, S., & Tarp, S. (2009). *Lexicography in the 21st Century: In honour of Henning Bergenholtz*. John Benjamins. <https://doi.org/10.1075/tlrp.12>
- Pete, T. (Ed.). (2011). *Medicine for the Salish language: English to Salish Translation Dictionary, Second Edition*. Salish Kootenai College Press. <https://www.nebraskapress.unl.edu/salish-kootenai-college-press/9781934594063/>
- Pettersen, B. E. (2011). Mind the digital gap: Questions and possible solutions for design of databases and information systems. In *Working with Traditional Knowledge: Communities, Institutions, Information Systems, Law and Ethics* (pp. 163–193).
- Pine, A., & Turin, M. (2018). Seeing the Heiltsuk orthography from font encoding through to Unicode: A case study using Convertextract. *Proceedings of the LREC 2018 Workshop “CCURL 2018 – Sustaining Knowledge Diversity in the Digital Age”*. <https://aidanpine.ca/static/cv/pdfs/Pine-Turin-Convertextract-2018.pdf>

- Rood, D. S., & Koontz, J. E. (2002). The Comparative Siouan Dictionary Project. In W. Frawley, K. C. Hill, & P. Munro (Eds.), *Making Dictionaries: Preserving Indigenous Languages of the Americas* (pp. 259–281). University of California Press.
- Rozen, D. L. (1985). *Place-names of the Island Halkomelem Indian people* (Master's thesis). University of British Columbia. <https://doi.org/10.14288/1.0096521>
- Sam, B. (2019). Mulyitul: Hul'q'umi'num' perspectives on a wedding ceremony (Master's project). Simon Fraser University.
- de Schryver, G.-m. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), 143–199. <https://doi.org/10.1093/ijl/16.2.143>
- Seward-Wilson, R. (2019). q'ushin'tul' 'u tunu shxw'a'luqw'a' (Master's project). Simon Fraser University.
- Stebbins, T. (2003). *Fighting Language Endangerment: Community Directed Research on Sm'algyax (Coast Tsimshian)*. Osaka Gakuin University. <https://doi.org/10.26826/1011>
- Stenzel, K. (2014). The pleasures and pitfalls of a 'participatory' documentation project: An experience in northwestern Amazonia. *Language Documentation & Conservation*, (8), 287–306. <https://scholarspace.manoa.hawaii.edu/handle/10125/24608>
- Stevens, M., Flinn, A., & Shepherd, E. (2010). New frameworks for community engagement in the archive sector: from handing over to handing on. *International Journal of Heritage Studies*, 16(1-2), 59–76. <https://doi.org/10.1080/13527250903441770>
- Suttles, W. (1990). Central Coast Salish. In W. Suttles & W. Sturtevant (Eds.), *Handbook of North American Indians, Volume 7: Northwest Coast* (pp. 453–457). Smithsonian.
- Svensén, B. (1993). *Practical Lexicography: Principles and Methods of Dictionary-making*. Oxford University Press.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press.
- Sylvester, C. (2019). lhalhukw' siiye'yu: An introduction to birds in the Hul'q'umi'num' world (Master's project). Simon Fraser University.
- Tarp, S. (2015). Structures in the communication between lexicographer and programmer: Database and interface. *Lexicographica*, 31(1), 217–246. <https://doi.org/10.1515/lexi-2015-0011>

- Tarp, S. (2019). A dangerous cocktail: Databases, information techniques and lack of vision. In *Studies on Multilingual Lexicography* (pp. 47–64). De Gruyter.
<https://doi.org/10.1515/9783110607659-004>
- Thompson, L. C., & Thompson, M. T. (Eds.). (1996). *Thompson River Salish Dictionary*. University of Montana Occasional Papers in Linguistics.
<https://doi.org/10.14288/1.0423560>
- Urbanczyk, S. C. (2004). Plurality and ablaut in Central Salish. In D. B. Gerdtts & L. Matthewson (Eds.), *Studies in Salish linguistics in honor of M. Dale Kinkade*. (pp. 429–453). University of Montana Occasional Papers in Linguistics.
- van der Velden, M. (2010). Design for the contact zone: Knowledge management software and the structures of indigenous knowledges. *Proceedings Cultural Attitudes Towards Communication and Technology*.
- van der Velden, M. (2013). Decentering Design: Wikipedia and Indigenous Knowledge. *International Journal of Human–Computer Interaction*, 29(4), 308–316.
<https://doi.org/10.1080/10447318.2013.765768>
- Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22(3), 233–243. <https://doi.org/10.1080/13645579.2018.1531228>
- Wattanawaroon, T. (2021). *Generalizing spreadsheet computation for evolving spreadsheets at scale* (A. Parameswaran, Ed.; Doctoral dissertation). University of Illinois at Urbana-Champaign. <https://www.ideals.illinois.edu/items/123247>
- Weber, T. (2021). The curation of language data as a distinct academic activity: A call to action for researchers, educators, funders, and policymakers. *Journal of Open Humanities Data*, 7. <https://doi.org/10.5334/johd.51>
- White, E. (Ed.). (1978). *Nanaimo Dictionary*. Nanaimo Band Council.
- Yong, H., & Peng, J. (2007). *Bilingual lexicography from a communicative perspective*. John Benjamins Publishing.
- Zhang, H., Alphonse, C., & Seymour, S. (2021). Creating an Online Hul'q'umi'num' Dictionary for Teachers and Learners [Conference presentation]. SAIL 2021, University of Arizona.

Appendix A

Hul'q'umi'num' Practical Orthography

	Practical Orthography	APA
1	'	ʔ
2	a	a
3	aa	a:
4	ch	č
5	ch'	č'
6	e	e
7	ee	e:
8	h	h
9	hw	x ^w
10	i	i
11	ii	i:
12	k	k
13	kw	k ^w
14	kw'	k' ^w
15	l	l
16	l'	l'
17	lh	ł
18	m	m
19	m'	m'
20	n	n
21	n'	n'
22	ou	u
23	oo	u:

	Practical Orthography	APA
24	p	p
25	p'	p'
26	q	q
27	q'	q'
28	qw	q ^w
29	qw'	q' ^w
30	s	s
31	sh	š
32	t	t
33	t'	t'
34	th	θ
35	tl'	θ'
36	ts	č
37	ts'	č'
38	tth	t ^θ
39	tth'	t' ^θ
40	u	ə
41	w	w
42	w'	w'
43	x	č
44	xw	č ^w
45	y	y
46	y'	y'

Appendix B

Hul'q'umi'num' Sound Inventory

The sound inventory of Hul'q'umi'num' is presented on the following pages, in both the practical orthography used by the HLCS and the Americanist Phonetic Alphabet (APA).

Hul'q'umi'num' Sound Inventory (Practical Orthography⁴⁴)

	Bilabial	Dental	Alveolar	Lateral	Palatal	Velar	Uvular	Glottal
						<i>unrounded</i>	<i>unrounded</i>	
						<i>rounded</i>	<i>rounded</i>	
Stop	<i>plain</i> p <i>glottalized</i> p'		t t'			k kw kw'	q q'	,
Affricate	<i>plain</i> tth <i>glottalized</i> tth'		ts ts'	tl'	ch ch'			
Fricative		th	s	lh	sh	hw	x	h
Resonant*	<i>plain</i> m <i>glottalized</i> m'		n n'	l l'	y y'	w w'		

*Resonants include nasals, liquids, and glides.

	Front	Central	Back
High	<i>short</i> i <i>long</i> ii		<i>short</i> ou <i>long</i> oo
Mid	<i>short</i> e <i>long</i> ee	u	
Low		<i>short</i> a <i>long</i> aa	

⁴⁴<https://sqwal.hwulmuhwqun.ca/learn/learning-the-sounds>

Hul'q'umi'num' Sound Inventory (APA⁴⁵)

	Bilabial	Dental	Alveolar	Lateral	Palatal	Velar	Uvular	Glottal
						<i>unrounded</i>	<i>unrounded</i>	
						<i>rounded</i>	<i>rounded</i>	
Stop	<i>plain</i> p <i>glottalized</i> p̰		t t̰			k kʷ k̰ k̰ʷ	q q̰ q̰ʷ	ʔ
Affricate		t ^θ t̰ ^θ	c c̰	ʃ ʃ̰	č č̰			
Fricative		θ	s	ʃ ʃ̰	š š̰	xʷ x̰ʷ	χ χ̰	h
Resonant*	<i>plain</i> m <i>glottalized</i> m̰		n n̰	l l̰ i ḭ	y y̰	w w̰		

*Resonants include nasals, liquids, and glides.

	Front	Central	Back
	<i>short</i> i <i>long</i> i:		<i>short</i> u <i>long</i> u:
High			
Mid	<i>short</i> e <i>long</i> e:	ə	
Low		<i>short</i> a <i>long</i> a:	

⁴⁵Americanist Phonetic Alphabet - https://en.wikipedia.org/wiki/Americanist_phonetic_notation

Appendix C

Full List of Dictionary Topics

The World Around Us

- Forest animals
- Salt water animals
 - » Mammals
 - » Fish
 - » Beach creatures
- Birds
 - » Bird words
 - » Songbirds — sqw'ulesh
 - » Big birds — thuthuws
 - » Water fowl — ma'uqw
- Creepy-crawlies
 - » Bugs
 - » Reptiles, amphibians, etc.
- Domesticated animals
 - » Pets
 - » Farm animals
 - » Animal sounds
- Plants
 - » Trees
 - » Other (grasses, shrubs)
 - » Berries
 - » Sea plants
- Nature
 - » Elements
 - » Celestial
 - » Geography of the land
 - » Geography of the ocean
 - » Weather

Us Humans

- Pronouns
 - » Independent pronouns
 - » Subject pronouns
- Body & Feelings
 - » Head & face
 - » Arms & legs
 - » Body core
 - » Internal organs
 - » Body actions
 - » Feelings & Senses
- Relations
 - » Parents
 - » Siblings & children
 - » Children
 - » Marriage, partners
 - » Extended family
- People
 - » People words
 - » Describing people

Life At Home

- Rooms of the house
 - » House
 - » Kitchen
 - » Living room
 - » Bedroom & blankets
 - » Bathroom
 - » Outside of the house
- Food – s’ulhtun
 - » Meals
 - » Fruit, vegetables & nuts
 - » Meat & animal products
 - » Drinks
 - » Other
- Making food
 - » Baking
 - » Cooking
 - » Dishes & setting the table
- Clothing & accessories
 - » Clothing
 - » Doing laundry
 - » Accessories
- Everyday items & tools
- Misc. cleaning & chores
- Personal grooming
- Childcare

People at Work & Play

- Occupations
 - » Jobs
 - » Farming
- Fishing
 - » What we use to fish with
 - » General fishing words
- Hunting
 - » What we use to hunt
 - » How & what we hunt
- Craftsmanship
 - » Spinning & weaving
 - » Carving & cutting
- Play
 - » Games
 - » Bonegame — sluhel'
 - » Ballgame — ts'uqwula'
 - » Card games
- Going to school — skwoukwul'
 - » Rooms & buildings
 - » Furniture
 - » School supplies

Out & About

- Places
 - » Buildings
 - » Place names
- Transportation, traveling, & shopping
 - » Transportation
 - » Traveling
- » Shopping
- Religion & Gatherings
 - » Religion
 - » Gatherings
 - » Longhouse

Time & Number

- Weeks, Months, & Seasons
 - » Week
 - » Month
 - » Season
- » Other
- Counting & amounts
 - » 0 through 100 (applied)
 - » Amounts

Some Language Basics

- Useful words
 - » Question words
 - » Location words
 - » Meeting & greeting
- Action Words
 - » Simple action words
 - » Verbs & progressives
- Describing words
 - » Size & shape
 - » Colours
 - » Other adjectives
 - » Adverbs
- Counting
 - » Counting with classifiers (1-5)
 - » Telling time

Appendix D

Transliterating between APA and the Practical Orthography

Transliteration will refer to the conversion of text in one orthography to another. Transliterating between APA and the practical orthography is useful for the students, especially as they advance their knowledge of linguistics and other Salish languages. It is also useful for students of the hə́nq̓əmiṇə́m (also called Downriver) dialect of Halkomelem, which uses APA phonetic characters. Hul'q'umi'num' and hə́nq̓əmiṇə́m are neighboring dialects, and language materials can be shared between the two.

There are two main MS Word macros that perform the transliteration task: one that transliterates APA into the practical orthography, and one that transliterates the practical orthography into APA.⁴⁶ In addition, there is a version of each macro that works only on a specific font, *BC Sans*. The files for all four of the transliteration macros can be found on Github at https://github.com/xyzhelen/hulq_word-macros.

The repository contains the following types of files:

1. (*.csv file) a list of ordered rules⁴⁷
2. (*.py file) a Python script which takes the rules and writes the word macro
3. (*.bas file) a MS Word macro written in VBA code

⁴⁶I was able to look at and modify code previously written by Zoey Peterson and Zack Gilkison.

⁴⁷These are **not** linguistic rules. For the most part they are ordered by string length so that the find & replace operations do not interfere with each other.

4. (*.docm file) a MS Word macro-enabled document which can be used to distribute the macro to the end user

For the font-specific macro, the user would first change the font of the Hul'q'umi'num' text to *BC Sans* to ensure that the transliteration rules would not apply to the English text. The script can be modified to specify any font⁴⁸ and *BC Sans* was only chosen because it (a) supported all the necessary APA characters and (b) was not a font that was used in existing documents.

A minimal working example of the VBA code for the *Practical orthography to APA* transliteration macro is shown in the following figure. Characters are referred to by the decimal base of their Unicode code point. It is advised to use this notation to write any special characters for the macro.

Each *find & replace* command corresponds to 5 lines of code in the macro, with examples shown in lines 12-16 and lines 19-23. Line 15 will find the string "tth" + "ChrW(8217)" which represents *tth'*; ChrW(8217) is the single right apostrophe character. Line 16 is the replacement text where "t" + ChrW(787) + ChrW(7615) represents *t^h*; ChrW(787) is the combining comma diacritic for glottalization, and ChrW(7615) is the raised theta.

⁴⁸See lines 8-9 in the code snippet.

Find & replace macro

```
1 Sub to_APA()  
2  
3 ' clear any existing formatting in the dialogue box  
4 Selection.Find.ClearFormatting  
5 Selection.Find.Replacement.ClearFormatting  
6  
7 ' specify the font as BC Sans (OPTIONAL)  
8 Selection.Find.Font.Name = "BC Sans"  
9 Selection.Find.Replacement.Font.Name = "BC Sans"  
10 Selection.Find.Format = True  
11  
12 '---start of find & replace operations  
13  
14 ' find & replace <tth'>  
15 With Selection.Find  
16     .Text = "tth" + ChrW(8217)  
17     .Replacement.Text = "t" + ChrW(787) + ChrW(7615)  
18 End With  
19 Selection.Find.Execute Replace:=wdReplaceAll  
20  
21 ' find & replace <tth>  
22 With Selection.Find  
23     .Text = "tth"  
24     .Replacement.Text = "t" + ChrW(7615)  
25 End With  
26 Selection.Find.Execute Replace:=wdReplaceAll  
27  
28 '--- end of find & replace operations  
29  
30 ' clear the dialogue box for the user  
31 Selection.Find.Execute Replace:=wdReplaceAll  
32 Selection.Find.ClearFormatting  
33 Selection.Find.Replacement.ClearFormatting  
34 Selection.Find.Text = ""  
35 Selection.Find.Replacement.Text = ""
```


Appendix E

Transcoding legacy fonts - Straight

Transcoding⁴⁹ will refer to the conversion of text between different encodings, in this case, a legacy non-Unicode encoding to the Unicode character encoding.⁵⁰ There are two legacy fonts that I encountered in legacy documents. Both fonts are named “Straight” and have the same glyphs, but different character mappings. I developed a MS Word macro that transcodes documents typed in either *Straight* font into a Unicode compliant font, BC Sans. The process is similar to the transliteration task, but the transcoder only works in one direction. The macro can be downloaded from Github at <https://github.com/xyzhelen/straight-font-to-unicode>.

Early versions of the Unicode standard, which was first released in 1991, did not contain all the characters necessary for typing Halkomelem in APA phonetics. This made it necessary for a custom font to be developed. In particular, the raised theta was not added until 2005.⁵¹

The original *Straight* font was developed in the mid 1990s by the linguist Charles Ulrich for Macintosh computers. A Windows version of the font with the same glyphs was later developed by a different person, and consists of different character mappings. To minimize confusion they will be referred to in this appendix as Straight (Mac) and Straight (Win) when necessary⁵². At the time they were developed, each font only worked on one operating system.

⁴⁹See Pine and Turin, 2018 for another example of font transcoding.

⁵⁰See Appendix D for transliteration between APA phonetics and the Hul’q’umi’num’ practical orthography.

⁵¹Modifier Letter Small Theta: <https://www.compart.com/en/unicode/U+1DBF>

⁵²To add to the confusion, keyboard layouts were also developed for each font. The keystrokes are the same but map to different characters.

While the *Straight* font proved extremely useful for many years, the continued use of this font would cause unnecessary compatibility issues. For example, one can not copy text in *Straight* and paste it elsewhere, and the *Straight* font has to be installed on any device to properly view a document that is typed in the *Straight* font.

It is necessary to transcode legacy material so that they can be read with any Unicode font. Transcoding legacy documents is a one-time process that helps preserve this important data.

Straight (Mac)

I used a free trial of FontLab 7⁵³ to analyze the font encoding and create the transcoder. The process for developing the transcoder macro was similar to the process described in Appendix D. Occasionally some symbols appear in documents which may result from previous versions of the transcoding process or font. For more information, see the Github at <https://github.com/xyzhelen/straight-font-to-unicode>.

Straight (Windows)

I came across relatively few documents that were typed in the Straight (Win) font. I was never able to get the Straight (Win) font to work properly on my computer and so the word macro only contains rules for characters that I could observe a few select documents. I compiled this table by comparing the same documents in .pdf and .doc file formats. The procedure to transcode the Straight (Win) font into Unicode is less straightforward. For details see the Github -https://github.com/xyzhelen/hulq_word-macros.

⁵³Font Lab - <https://www.fontlab.com/font-editor/fontlab/>

The following table shows only the 19 character correspondences that were needed to properly transcode those documents. The Unicode codepoints are written in decimal base.

Straight (Win)		Unicode	
Glyph	Unicode codepoint	Glyph	Unicode codepoint
š	(353)	§	(167)
č	c + (780)	Æ	(198)
ě	“k” + (787)	û	(251)
đ	q + (787)	ï	(207)
ž	(411) + (787)	Ã	(195)
š	x + (780)	Å	(197)
č	“c” + (787)	□	(141)
ě	“m” + (787)	μ	(181)
đ	“n” + (787)	◊	(186)
š	“l” + (787)	Â	(194)
đ	“y” + (787)	´	(180)
š	“w” + (787)	·	(183)
?	(660)	Ö	(214)
ł	(322)	¨	(168)
ə	(601)	;	(59)
đ	“p” + (787)	¹	(185)
θ	(952)	Ä	(196)
·	(183)	¥	(165)
w	(695)	¿	(191)