Woods Cree (*nîhithawîwin*): A Grammatical Description and Computational Modelling

by

Daniel Benedict Dacanay

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science

Department of Linguistics
University of Alberta

Abstract

In writing this thesis, I aim to accomplish three primary objectives, all of which are intended to increase the quality and quantity of language resources available in Woods Cree (*nîhithawîwin*, ISO:cwd). Firstly, I provide here a thorough descriptive overview of the Woods Cree phonological and morphological systems, describing in detail their mutual interactions and variations across geographic and cross-generational lines. Secondly, I describe the compilation of an online, bilingual Woods Cree-English corpus, containing tens of thousands of tokens from geographically, diachronically, and stylistically diverse sources, making use of this corpus throughout to evaluate various phenomena in the language. Thirdly, I describe the synthesis of my metalinguistic findings into a computational model of Woods Cree morphology, capable of both recognising and generating inflected wordforms in the language. In turn, this model was used to create a morphologically intelligent online dictionary of Woods Cree (https://itwiwina.altlab.dev/), adapted from existing computational tools created for the related dialect of Plains Cree.

kâ-mâmiskôcikâtîki ôma masinahikanihk

nikî-masinahîn ôma masinahikan ta-tôtamân nisto kîkwaya, osâm mwâc î-tîpipathik mîtho-nîhithawîw-âpacihcikana ihtakwahki anohc. nistam, ôma masinahikanihk nikiskinwahîn nîhithawi-pîkiskwîwin, tânisi îtihtâkwahk, tânisi itwîwina ikwa mîna pîkiskwîwina î-osîhtâniwahki. nikiskinwahîn mîna tânisi nîhithawîwin pâh-pîtos îtihtâkwahk tahto iskonikanihk, ikwa tânisi osk-âyak ikwa kîhti-ayak pâh-pîtos pîkiskwîwak. akwani, nikiskinwahîn tânisi î-mâwasakonamân nanâtohk nîhithawasinahikanisa pâh-pîtos iskonikanihk ikwa otînâsihk ohci. nikî-âpacihtân ôho nîhithawasinahikanisa nawac ta-pî-kiskîthihtamân tânisi ôma nîhithawi-pîkiskwîwin î-atoskîmakahk. pîthisk nikî-âpacihtân kahkithaw ôho kîkwaya ta-osihtâyân mamahtaw-âpacihcikanihkân kâ-nisitohtahk nîhithawîwin ikwa mîna kwayask ka-kî-masinaham nîhithaw-itwîwina. kika-kî-miskîn ôma nîhithawi-mamahtaw-âpacihcikanihkân ôta (https://itwiwina.altlab.dev/)*, ôma nîhithaw-itwîwasinahikanihk.*

Preface

This thesis is a wholly original work by Daniel Benedict Dacanay. The research belying this thesis was funded in part by the SSHRC partnership grant '21st Century Tools for Indigenous Languages [21C]' (#895-2019-1012), operated by the Alberta Language Technology Lab (ALT Lab) at the University of Alberta.

Certain aspects of the computational modelling also benefited from the previously existing work within the 21C project. The computational morphological model for Woods Cree presented in this thesis adapted the phonological rules and the affixal morphology for nouns and verbs from the revised version of the Plains Cree model developed by Atticus Harrigan, Katherine Schmirler, and Antti Arppe.

ohci nîhithawak, kayâs, anohc, mîna kâkikî
To the Cree people, then, now, and forever

Acknowledgements
To begin, I bow in reverence to the many researchers, speakers, elders, and oft-thankless enthusiasts who have written in and of the Woods Cree language before me; it is on their shoulders that this thesis and the fruits of it stand, and it is to them that I am most inimitably grateful.

After these, I attribute due credit to my supervisor, Dr. Antti Arppe, although my attempts will doubtless fall much short of that which is truly owed. Having been the principal guide of my scholarly pursuits for over a fifth of my life, there is not a scrap of research which I have performed, either here or elsewhere, which does not bear either his influence or his direct command. Although it would be hyperbole to say that I owe everything of my academic career to him, it would be hyperbole only by the width of a hair. To him, I say *olen vertaansa vailla kiitollinen*.

I would like also to give my thanks to Dr. Arok Elessar Wolvengrey and Professor Solomon Ratt (SOM) of the First Nations University of Canada, both of whom have contributed greatly to the descriptive portions of this thesis. To the former, for his peerless insight on the intricacies of Western Cree morphosyntax, I say *le fael*. To the latter, for his decades of work proliferating and promoting his native language (even now, in his well-earned retirement), I say *kinanâskomitin.*

I must also thank Rose Makinaw of the Ermineskin Cree Nation, whose patience with me (and my interminable silly questions) have been an invaluable aid in my understanding and research of the Cree language over the years, and whose contributions to this work in particular, and to the cause of language revitalisation in general, are greatly lauded and appreciated. To her, I say *kâkikê ê-mâh-miywêyihtamihiyan*.

I must also express my gratitude to the Lac La Ronge Indian Band for granting me permission to use their dictionary as a basis for this thesis; it is my sincere hope that the fruits of my research should prove advantageous to them and their efforts at preserving their remarkable language. To them, I say *âhkamîthimok*

Finally, I express my acknowledgement to the Social Sciences and Humanities Research Council of Canada, whose funding through the 21st Century Tools for Indigenous Languages Partnership was instrumental in carrying out the research described here, and to my colleagues at the University of Alberta, most notably Maaz Ahmad Shaikh, for their help in reviewing this work.

Of course, any errors of fact or interpretation are solely of my own making.

Daniel Benedict Dacanay,
sed soli Deo gloria

TABLE OF CONTENTS

List of Tables

List of Figures

Interlinear Gloss Abbreviations

[]	Contents of brackets constitute a lexemic or affixal unit (stem or affix)	LV	Linking vowel
-	Affix boundary	MED	Medial
–	Orthographic hyphen (in interlinear glosses)	NA	Animate noun
0	Impersonal subject	NC	Not in Corpus, constructed sentence
0`	Impersonal obviative subject	NDA	Dependent Animate Noun
1PL	First-person plural exclusive	NDI	Dependent Inanimate Noun
1SG	First-person singular	NEG	Negation
12PL	First-person plural inclusive	NI	Inanimate noun
2PL	Second-person plural	OBV	Obviative
2SG	Second-person singular	OBJ	Object
3PL	Third-person plural	ORD	Ordinal
3SG	Third-person singular	PFS	Particle-forming suffix
3`	Third-person obviative	PK	Pukatawagan, Manitoba
3`PL	Third-person plural obviative (VII only)	PL	Plural
3``	Third-person further obviative	PN	Pelican Narrows, Saskatchewan
ABST	Absentative	POSS	Possession
ADD.FOC	Additive focal (pronoun)	PrA	Proto-Algonquian
ANIM	Animate	PRET	Preterit
BR	Brochet, Manitoba	PRON	Pronoun
CNJ	Conjunct	PROX	Proximal

COM	Comitative
DEM	Demonstrative
DIM	Diminutive
DSTR	Distributive
DUB	Dubitative
FIN	Final
FOC	Focus marker
FUT	Future
INAN	Inanimate Actor (on verbs); Inanimate (elsewhere)
INIT	Initial
IND	Independent
INTERR	Interrogative
KN	Kinoosao, Saskatchewan
LC	Linking consonant
LOC	Locative
LR	La Ronge, Saskatchewan
PRSP	Prospective aspect
PST	Past tense
RDPLS	Strong reduplication
RDPLW	Weak reduplication
REL	Relational
SG	Singular
SL	South Indian Lake, Manitoba
SM	Stanley Mission
SUBJ	Subjunctive
VAI	Animate Intransitive verb
VII	Inanimate Intransitive verb
VOC	Vocative
VTA	Transitive Animate verb
VTI	Transitive Inanimate verb
UL	Unknown Location

Glossing Format

The numbered interlinear translations provided throughout this thesis are formatted according to the Leipzig Glossing Rules (Comrie et al. 2008), albeit with some modifications for the particularities of the data presented. Line by line, the formatting is as follows. To begin, there is an optional underlined heading, which serves to explain the feature being highlighted in a given example:

> VTA Conjunct Morphology from the Third Person Inverse Set

If the example stems from a historical text, there is then a line of non-italicised text representing the original, historical orthography. Since not all examples stem from historical sources, this line is optional:

> Kéche-mánneto otáweemayoo únnehee gà sàkehíkoot

After this, the example is written in the Standard Roman Orthography (see Section 2.1.1); this line is always present, and is italicised:

> *kisî-manitow ohtawîmîw anihi kâ-sâkihikot*

After this comes a transcription of the example in the International Phonetic Alphabet, provided such a transcription is relevant. If it is not, this line is excluded:

> [ˌkɪsiːˈmʌnˌtow ohˈtawiːˌmiːw ˈʌnhɪ ˌkaːsaːˈgihiˌkʊt]

Next is a representation of the example with morpheme boundaries, contractions, and other morphophonological reductions explicitly indicated; again, this line is optional, depending on the nature of the example. If this boundary marked line is present, then the following line provides an English translation of it on the morphemic level:

> kisî-manitow ohtawîm-îw anihi
> God be.father(VTA)-3SG>3`.IND DEM.MED.ANIM.OBV
>
> kâ-sâkih-ikot
> CNJ-love(VTA)- 3`>3SG.CNJ

The next line is a prosaic, ‘free’ translation of the example. This translation line is also mandatory. The ‘free’ translation line also includes a geographic tag in parentheses (the tags themselves being listed in the above glossary), as well as an in-text citation if the example stems from a source other than the text corpus (see Section 1.6):

‘God is the father of those who love Him’ (UL) (Howse 1844:264)

Finally, an optional line for notes may be inserted underneath the free translation, if further contextual information is required:

Note: Given by Howse as an example of a Cree translation of the English passive

In all, the only two lines which are obligatorily present in all examples are the SRO transcription and the free translation; however, most will have at least one additional line depending on the nature of the example.

Heading
Historical Orthography
Standard Roman Orthography
[IPA transcription]
Underlying morphemic representation
Interlinear gloss translation
‘Free translation’ (Geographic origin) (Citation [if not in corpus])
Note: Additional relevant information

VTA Conjunct Morphology from the Third Person Inverse Set
Kéche-mánneto otáweemayoo únnehee gà sàkehíkoot
kisî-manitow ohtawîmîw anihi kâ-sâkihikot
[ˌkɪsiːˈmʌnˌtow ohˈtawiːˌmiːw ˈʌnhɪ ˌkaːsaːˈgihiˌkʊt]
kisî-manitow ohtawîm-îw anihi
God be.father(VTA)-3SG>3`.IND DEM.MED.ANIM.OBV

kâ-sâkih-ikot
CNJ-love(VTA)- 3`>3SG.CNJ
‘God is the father of those who love Him’ (UL) (Howse 1844:264)
Note: Given by Howse as an example of a Cree translation of the English passive

FST Tag Abbreviations

`>`	Morpheme boundary	`+Fut+Cond`	Subjunctive (‘Future-Conditional’)
`+0Sg/Pl`	Inanimate Actor (INAN)	`+I`	Inanimate
`+1Pl`	First-person plural exclusive (1PL) subject	`+II`	Inanimate Intransitive
`+1PlO`	First-person plural exclusive (1PL) object	`+Incl`	Additive Focal (‘Inclusive’)
`+1Sg`	First-person singular (1SG) subject	`+Ind`	Independent
`+1SgO`	First-person singular (1SG) object	`+Ipc`	Indeclinable Particle
`+12Pl`	First-person plural inclusive (12PL) subject	`+Loc`	Locative
`+12PlO`	First-person plural inclusive (12PL) object	`+Med`	Medial
`+2Pl`	Second-person plural (2PL) subject	`+N`	Noun
`+2PlO`	Second-person plural (2PL) object	`+Obv`	Obviative
`+2Sg`	Second-person singular (2SG) subject	`+TA`	Transitive Animate
`+2SgO`	Second-person singular (2SG) object	`+Pers`	Personal (pronoun)
`+3Pl`	Third-person plural (3PL) subject	`+Pl`	Plural
`+3PlO`	Third-person plural (3PL) object	`+Pron`	Pronoun
`+3Sg`	Third-person singular (3SG) subject	`+Prox`	Proximal

+3SgO	Third-person singular (3SG) object	PN/	(Lexical) Prenoun
+4Sg/Pl	Third-person obviative (3`) subject	PV/	(Grammatical) Preverb
+4Sg	Third-person singular obviative (3`) subject (VII only)	+Px1Sg	First-person singular (1SG) Possessor
+4Pl	Third-person plural obviative (3`) subject (VII only)	+Px2Sg	Second-person singular (2SG) Possessor
+4Sg/PlO	Third-person obviative (3`) subject object	+Px3Sg	Third-person singular (3SG) Possessor
+5Sg/Pl	Third-person further (3``) obviative subject	+Px1Pl	First-person plural exclusive (1PL) Possessor
+5Sg/PlO	Third-person further (3``) obviative object	+Px12Pl	First-person plural inclusive (12PL) Possessor
+A	Animate	+Px2Pl	Second-person plural (2PL) Possessor
+AI	Animate Intransitive	+Px3Pl	Third-person plural (3PL) Possessor
+Cnj	Conjunct	+Px4Sg/Pl	3` Possessor
+D	Dependent	+PxX	Unspecified Possessor
+Dem	Demonstrative	+TI	Transitive Inanimate
+Der/Dim	Diminutive	+V	Verb
+Dist	Distal	+Var/East	Eastern Variant
+Distr	Distributive	+X	Unspecified Actor
+Err	Nonstandard Affixation (‘Error’)		

Introduction

Among the many challenges faced by the Indigenous languages of Canada in the 21st century, two of the foremost are a lack of descriptive documentation and a lack of freely available language resources. Without the former, the development of critical language learning tools such as school curricula and dictionaries is near impossible, and without the latter, communities lack the necessary tools to mount an organised program of language revitalisation. These problems, which have existed in Indigenous language communities for many decades, have become increasingly apparent in the emergent social gathering place of the internet. In recent years, it has become common for segments of all populations, including the Indigenous, to spend a substantial portion of their social, professional, and personal lives on the internet. As such, facilitating the online presence and usage of Indigenous languages is a top priority for any contemporary language revitalisation effort.

This thesis therefore attempts to address the lack of online language resources for a particular Indigenous language; Woods Cree, a Central Algonquian language spoken throughout Saskatchewan and Western Manitoba, as well as to fill various gaps in the documentation of the language's grammar. I accomplish this in three distinct sections. In the first section, I provide an overview of the Woods Cree language, focusing on its demographics, geographic distribution, relation to neighbouring languages, dialectal variation within the language, and existing documentation. I also describe the creation of a Woods Cree corpus, which I will make freely available online.

In the second section, I provide a descriptive account of Woods Cree phonology and morphosyntax. This section largely builds off of the documentary field work of Donna Starks (1992) and Jennifer Greensmith (1985a), as well as descriptive accounts of related Cree varieties written by H.C. Wolfart (1973), Arok Wolvengrey (2011a), and C. Douglas Ellis (2000; 2004; 2016). The data belying this descriptive section stems primarily from the corpus described in the previous section. Using this, I illustrate dialectal variation within Woods Cree, as well as general typological trends between Woods Cree and neighbouring Cree varieties.

In the final section, I describe the creation of a finite-state transducer-based morphological model of Woods Cree, based on an earlier model designed for Plains Cree (Snoek et al. 2014; Harrigan et al. 2017). This model is able to both generate and recognise full paradigms for Woods Cree nouns and verbs, as well as to provide analyses for these forms in running text. Using this model, I also demonstrate an intelligent online dictionary of Woods Cree, which can both generate paradigms automatically for each entry as well as accept and analyse inflected forms as search queries. Finally, I discuss the applications and limitations of this technology, as well as future avenues for research into Woods Cree language technology.

1. BACKGROUND AND OVERVIEW OF WOODS CREE

In this section, I will provide a general introduction to the Woods Cree language, discussing its geographic distribution, its relation to other languages in the Cree dialect continuum, and its present-day demographics and vitality. In addition, I will introduce various aspects of subdialectal variation and establish a rough geographic and linguistic delineation between Eastern and Western varieties. After this, I will provide an overview of existing documentation and literature in Woods Cree, contrasting with other related languages. Finally, I will describe the construction and refinement of a bilingual text corpus and an emergent audio corpus for Woods Cree which will serve as key resources in later sections.

1.1 Geographic Distribution

Woods Cree (ISO:cwd, *nîhithawîwin* (see Appendix A: Nomenclature)) is a Central Algonquian language spoken primarily in the forested northern regions of Saskatchewan and Manitoba, as well as to a lesser extent in northeastern Alberta (Wolvengrey 2011a:8). As its name implies, it is one of the nine major dialect members of the Cree-Montagnais-Naskapi dialect continuum, a geographically expansive dialect chain stretching from the Rocky Mountains in British Columbia to the Atlantic coast of Labrador. More specifically, it is a part of the Western sub-branch of this dialect continuum, which extends roughly as far east as the border between Ontario and Québec (see Figure 1). Although rigid dialectal boundaries within this continuum are difficult to define, Woods Cree is primarily spoken in the region between the Churchill and Nelson River, north of Lac La Ronge (Wolfart 1973:10). Within the dialect continuum, this situates Woods Cree to the northeast of Plains Cree (ISO:crk, *nêhiyawêwin*) and northwest of Swampy Cree (ISO:csw, *ininîmowin*), two dialects to which it is closely related:

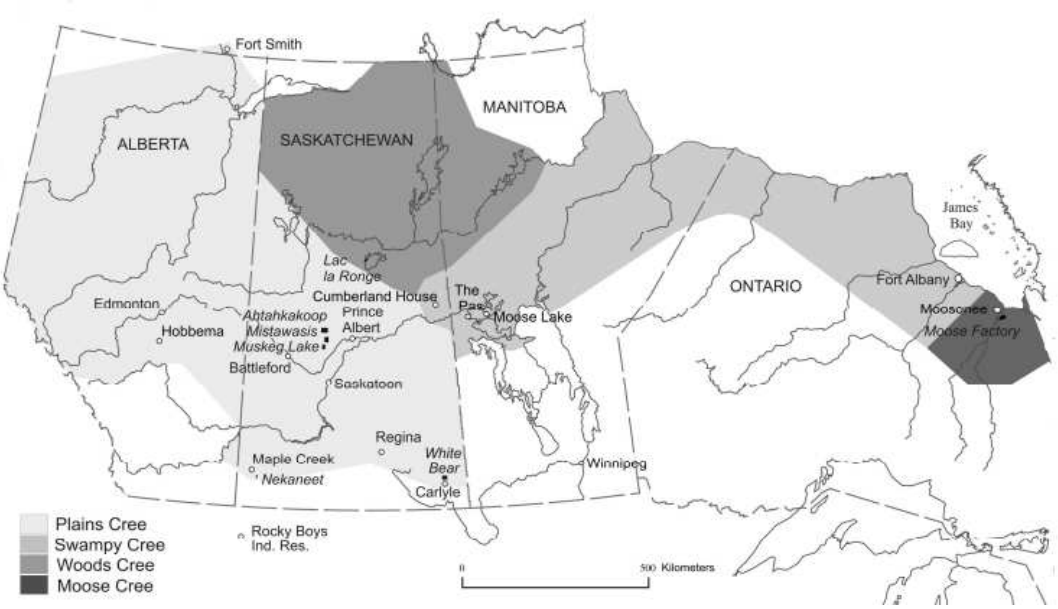

Figure 1. The Western sub-branch of the Cree dialect continuum (Wolvengrey 2011a)

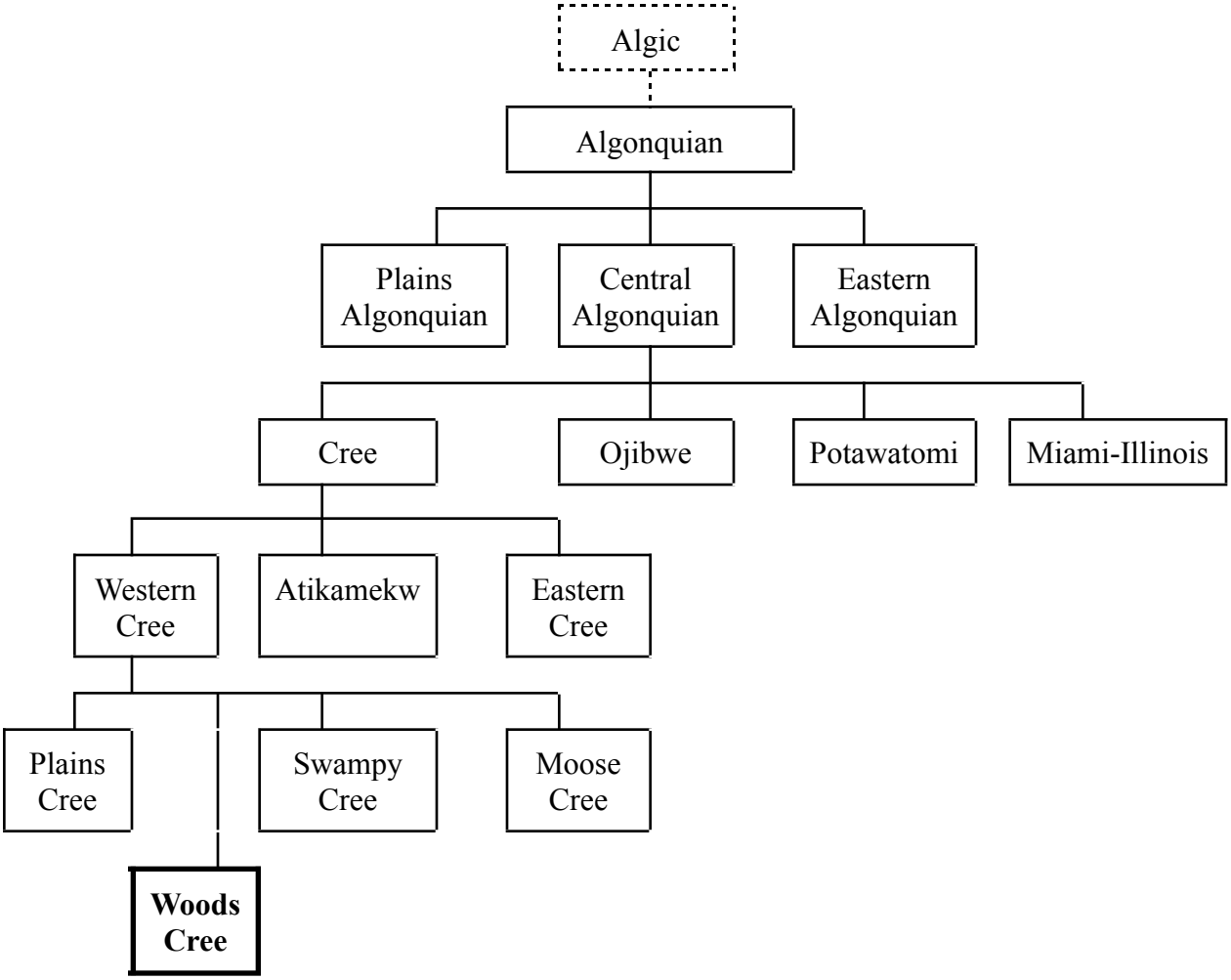

Figure 2. A simplified Algonquian family tree, adapted from Mii Dash Geget (2023) and Wolvengrey (2011a)

On account of their traditionally nomadic lifestyle (described in detail in the works of Robert Brightman (1993) and James Smith (1981;1987)), the geographic distribution of Woods Cree overlaps considerably with those of its neighbouring dialects. Speakers and communities near the dialect boundaries with Plains and Swampy Cree are often bidialectal, or combine features from multiple dialects; this is discussed in further detail in Section 1.3. As recently as the 19th century, Woods Cree was spoken across eastern Manitoba and Ontario, possibly as far east as Québec (Pentland 1978:124); however, these areas are now uniformly populated by Swampy Cree, Moose Cree, and East Cree communities. Due to modern urban migration, small pockets of Woods Cree speakers may also be found in cities throughout the Canadian prairies, particularly in Saskatchewan (Rock Cree Language Council 2023).

1.2 Demographics

In contrast to their historical nomadism, most modern Woods Cree speakers live in permanent settlements and reserves. Owing to their traditional means of subsistence, which are largely dependent on hunting, trapping, and fishing, as well as the relatively low biomass of the northern boreal forests in which they are situated[1], these settlements tend to be sparsely dispersed and geographically isolated. This pattern of relative geographic remoteness has led to difficulties in collecting thorough statistical enumerations of any kind in most Woods Cree communities, including those on Woods Cree language usage. This is exacerbated by the tendency of Woods Cree speakers (and, indeed, Cree speakers at large) to refer to their own language (at least, in English) simply as 'Cree', rather than a specific dialect of Cree (see Appendix A); for example, of the total 74,580 individuals who identified their L1 as 'Cree languages' in the 2021 Canadian Census, 49,865 list their mother tongue simply as 'Cree (n.o.s)", with only 24,715 listing their specific dialect.

Nonetheless, extrapolations based on census data can provide a rough estimate of the total speaker count. Of the 24,715 Cree L1 speakers on the national census who did specify their dialect, 3,705 (or, 14.9%) list "Nihithiwiwin (Woods Cree)" as their mother tongue. Applying this proportion to the 49,865 speakers who did not list their dialect group and adding this number to the explicitly recorded count, one may extrapolate the total number of Woods Cree speakers at 11,180, a number which roughly corresponds to the higher end of the 2015 Ethnologue estimate of "less than ten thousand", as well as with Jennifer Greensmith's 1985 estimate of 8,000 speakers (Greensmith 1985:2). Based on these numbers, a broad estimation of slightly more or slightly less than 10,000 speakers may be posited. The precise geographic concentration of these speakers within the Woods Cree dialect region is not known.

1.2.1 Vitality

Like most Indigenous languages in Canada, and indeed most members of the Cree dialect continuum, Woods Cree is faced with various aspects of language endangerment, most notably an unsustainably low level of intergenerational transmission and the displacement of Woods Cree in favour of English in most societal domains (particularly those related to government and mass media). Although knowledge of the language is widespread in most communities among adults above 30, increasing contact with the English language through exposure to modern media, combined with strong economic and social pressures from the wider world towards the use of English, have left much of the younger generation with either a non-fluent or non-existent grasp of Woods Cree. This decline was severely exacerbated in the 20th century through the Indian Residential School system, a mandate from the Canadian federal government to educate Indigenous children away from their communities and exclusively in English or French, with any use Indigenous languages by the (often monolingual) children being violently punished. This

[1] Leighton (1982) provides a detailed account of the Woods Cree ecological landscape

aggressively enforced educational isolation is often credited by speakers as being the chief reason behind the language's decline (e.g. Ratt 2023).

Although it is difficult to ascertain precise demographic statistics for Woods Cree as a whole, there does exist survey information on language vitality in specific communities. Most notably, in 2021, the Rock Cree Language Council, a Woods Cree language advocacy group, performed a sociolinguistic survey of 7 communities in the Peter Ballantyne Cree Nation, a large Woods Cree-speaking First Nation in remote northern Saskatchewan, in which they deemed the language "threatened" in all communities surveyed.

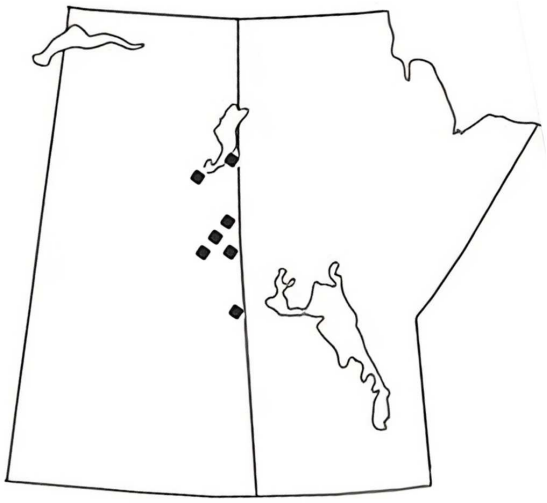

Figure 3. Location of surveyed communities in the Peter Ballantyne Cree Nation in Saskatchewan and Manitoba

Across the survey, 87.5% of the total 1704 respondents self-reported as Woods Cree L1 speakers, with 96.7% of respondents above 50, 88.9% of respondents between 30 and 49, 78.9% of respondents between 17-29, and 72.4% of respondents younger than 16 identifying as such. However, self-reported fluency was generally much lower among younger speakers; 88.6% of respondents above the age of 50 reported their proficiency as 'fluent', compared to 66.1% of respondents aged 30-49, 26.4% of respondents aged 17-29, and only 3% of respondents younger than 16. By comparison, 10.9% of respondents above 50 reported some fluency, but not complete fluency, as did 30.9% of respondents between 30-49, 66.2% of respondents between 17-29, and 81.2% of respondents 16 or below. 81.6% of respondent households reported that Woods Cree was regularly used as a home language, but in a substantial majority of these cases, both Cree and English were used as home languages. Only 7.6% of households reported Woods Cree as the sole home language.

	Number of Respondents	Age 6-16		Age 17-29		Age 30-49		Age 50+	
Woods Cree as L1		Yes	No	Yes	No	Yes	No	Yes	No
Deschambeault Lake	384	14	10	62	17	161	8	109	2
Kinoosao	6	0	0	1		1	1	3	0
Pelican Narrows	669	50	4	171	3	258	0	181	0
Sandy Bay	295	5	8	24	42	59	44	96	14
Southend	288	27	5	62	14	76	8	102	1
Sturgeon Landing	18	0	2	2	3	5	3	3	0
Denare Beach	44	1	8	1	7	9	7	9	0
Total		97	37	323	86	569	71	503	17

Table 1. Number of survey respondents reporting Woods Cree as their L1 in the Peter Ballantyne Cree Nation

	Number of Respondents	Age 6-16			Age 17-29			Age 30-49			Age 50+		
Fluency Level		High	Semi.	Non.	High	Semi.	Non.	High	Semi.	Non.	High	Semi.	Non.
Deschambeault Lake	384	0	22	2	14	64	1	123	47	0	102	6	1
Kinoosao	6	0	0	0	0	1	0	0	2	0	3	0	0
Pelican Narrows	669	4	49	0	89	83	1	231	29	0	174	7	0
Sandy Bay	295	0	9	4	0	46	20	7	82	15	69	40	1
Southend	288	0	21	12	5	67	4	61	21	2	94	1	0
Sturgeon Landing	18	0	2	0	0	3	2	2	6	0	3	0	0
Denare Beach	42	0	5	3	0	7	2	2	12	2	7	2	0
Total		4	108	21	108	271	30	426	199	19	452	56	2

Table 2. Self-reported fluency of Woods Cree-speaking survey respondents in the Peter Ballantyne Cree Nation. 'High' fluency indicates that the respondent identified with the survey category 'fluent', 'Semi.' fluency indicates that the respondent identified with one of the following statements on fluency: "I can make myself understood, but have some problems with it", "I know a lot of words and phrases, but have a hard time communicating", or "I know some vocabulary, but can't speak in sentences. "Non." fluency indicates that the respondent self-identified as a Woods Cree speaker, but also identified themselves as "Not at all" fluent.

One of the only aspects of language use which did not differ between age groups was the perceived importance of the language: when posed with the question "What is the importance of the Cree language to you?", 93.1% of total respondents answered "Very Important", the maximum rating offered.

Outside of this internal survey, external prognoses of Woods Cree's vitality have drawn similar, albeit much less detailed, conclusions. For example, an evaluation performed by the UNESCO *World Atlas of Languages in Danger* described Woods Cree, as well as its neighbouring dialects of Plains and Swampy Cree, as "Vulnerable", the second highest rank (after "Safe") on the five-point UNESCO *Language Vitality and Endangerment Scale*. Ethnologue ranked the language as "Endangered", the second lowest ranking on its four-point scale of language vitality, with the description that it is "no longer the norm that children learn and use this language". Ethnologue also noted the status of resource development in the language to be "Still", their lowest possible rank.

As a complement to these quantitative accounts of the vitality of Woods Cree, I will also include here a brief discussion of the language's shifting status from a more qualitative, community-based perspective, namely, that of Ida Swan, a native speaker from Pelican Narrows, Saskatchewan. In her 2001 masters thesis *Language Shift: A Study of Three Generations within a Cree Family*, Swan provides a detailed overview the language usage habits of individuals from multiple generations of her own family, highlighting the contrasts between her parental generation, largely monolingual Woods Cree speakers, her own generation, who are fluently bilingual in Woods Cree and English, and the generation of her children, many of whom are non-fluent heritage speakers. Speaking of her father, Angus Merasty (born 1923), whom she states to be broadly representative of speakers of his generation, Swan provides this description:

> "My father speaks Cree almost exclusively … The only time my father speaks English is when he is forced to communicate with non-Cree speakers when no one is around to translate. He communicates in the English language when he goes shopping and when he has to see a doctor. Recently, more and more English is being spoken in my parents house but normally when this happens, my father will not get involved in the conversation. If he does, he will converse in Cree only." (Swan 2001:74)

As a representative of her own generation, Swan gives the example of her younger sister, Margaret Brass (born at some point in the 1950s or 1960s). Swan describes Margaret as "an excellent Cree speaker" who speaks both Cree and English on a daily basis. Margaret learned English through the residential school system, and reported that her schoolmates at the time "preferred to speak in English", and that knowledge of English was necessary at the school to "get along with other people". Although a fluent speaker, Swan notes that Margaret's speech shows "undue influence" from English, with frequent, phrase-length code-switches into English.

> "*ikwa nipîtawahnan î-mâci-pâh-pakamawât* it's an old, an old tree stump that's dead*, î-nipit isa ana* tree *ana isa"*
> "and so they came to start hitting on it, it's an old, an old tree stump that's dead, it was apparently dead, this tree, this one" (Swan 2001:177)

As representatives of her children's generation, Swan uses a number of examples, beginning with her own daughter, Carla Swan (born 1976). Carla is described as a fluent speaker, but with considerable influences in her speech both from Plains Cree (the dialect taught in her high school) and English. Although fluent, the domains in which Carla is able to speak Cree are considerably fewer than those of her parents, limited almost exclusively to the homes of her various family members. Swan also describes her sister Margaret's five children, of whom only the oldest speaks primarily in Cree, with the two middle children preferring English, and the two youngest children speaking both roughly equally. Of the younger members of this generation, Swan includes the following direct quote from her father:

> "I am not pleased with them. For example, Candi, we speak Cree to her all the time. She will not speak Cree. She speaks English most of the time. And my grandchildren, the ones who go to school, I don't know what they are saying because I don't speak English" (Swan 2001:34)

Swan also discusses a visit which she made in February 2000 to a school in Pelican Narrows, where she noted students "using the English language as well as Cree, to speak to their friends, while others were using only English". Of note, she remarks that "nowhere did [she] hear a conversation between young people spoken completely in the Cree language" (Swan 2001:10).

As demonstrated both by the statistical enumerations of the Rock Cree Language Council and the personal accounts of Ida Swan, there is a general consensus among contemporary authors and community members that the Woods Cree language is in a definite state of decline. Although not in immediate danger of extinction (as many Canadian Indigenous languages are), Woods Cree is in clear peril of being displaced by English in most communities within one or two generations. Particularly among the youth, most daily communication either is, or can be, accomplished solely in English or in a mixture of English and Cree. Few domains of regular language use remain in which solely Woods Cree is acceptable.

1.3 Relation to Other Dialects

To date, relatively little research has been conducted on the topic of mutual intelligibility between Cree dialects (MacKenzie 1980; Pentland 1978; Gordon 1965). However, the general consensus among linguists is that the Cree dialect continuum may be split into two broad geographic zones, these being the Western Cree zone (encompassing Plains Cree, Woods Cree,

Swampy Cree, and (sometimes) Moose Cree) and the Eastern Cree zone (encompassing East Cree, Innu, Naskapi, and (sometimes) Atikamekw), with the boundary between the two zones existing either on the eastern edge of the Swampy Cree dialect area (Wolvengrey 2011a:8) or at the Atikamekw dialect area (MacKenzie 1992:274; Wilson 2022:2). Whether Moose Cree and Atikamekw are to be considered part of the Western Cree group is thus a matter of debate; however, in this investigation, I will include only the former, defining Western Cree as encompassing Plains, Woods, Swampy, and Moose Cree, with Atikamekw as its own distinct linguistic entity. However, in practice, I will primarily be comparing language data from Woods Cree's immediate neighbours, Plains Cree and Swampy Cree.

The chief differences between the Western Cree dialects are phonological and lexical, consisting of regular, predictable sound shifts and the preference or dispreference of certain lexical items. Although some morphosyntactic differences do exist between (and within) these Cree dialects, these differences seldom seem to pose an obstacle to interdialectal intelligibility. The most immediately conspicuous difference between Woods Cree and its neighbouring dialects is its use of the phoneme /ð/ as a reflex of the Proto-Algonquian *r[2] (Goddard 1994). By contrast, Plains Cree (or 'y-dialect') uses /j/ (⟨y⟩) for this reflex, Swampy Cree (or 'n-dialect') uses /n/ (⟨n⟩), and Moose Cree (or 'l-dialect') uses /l/ (⟨l⟩). For example:

1) Proto-Algonquian *ni·ra[3] ('1sg pronoun')
   Plains Cree: *niya*
   Woods Cree: *nîtha*
   Swampy Cree: *nîna*
   Moose Cree: *nîla*

This reflex is the primary isoglossic feature used to distinguish between Western Cree dialects. The Woods Cree dialect area is therefore typically defined linguistically as encompassing all communities in which /ð/ predominates as the reflex of Proto-Algonquian *r. However, in applying this definition, it should be kept in mind that dialectal multilingualism among Woods Cree speakers is common, particularly among the youth (i.e. Swan 2001:79), as is interpersonal mobility between Cree settlements of varying dialects. In the Rock Cree Language Council survey, for instance, all but one community (Sturgeon Landing, with a population of 111) reported a minority of respondents speaking Plains Cree and/or Swampy Cree. Dialectal code-switching is also common. Consider the following example, where a bilingual Woods Cree L1 speaker reading from a text in Plains Cree pronounces some words (underlined) in their Woods Cree form and others (bolded) in their Plains Cree form:

[2] There is some controversy surrounding the exact nature of this segment in Proto-Algonquian; notably Bloomfield (1925;1946) reconstructed this segment as /l/.
[3] All Proto-Algonquian reconstructions are taken from https://protoalgonquian.atlas-ling.ca/, which in turn stem from Hewson (1993)

2) *iyako wiya sîmâk kiskîthihtam î-nipahtâkîyit omisa*

iyako	**wiya**	sîmâk	kiskîthiht-am	î-nipahtâkî-**yit**
PRON.INAN.SG	FOC	immediately	know(VTI)-3SG.IND	CNJ-kill(VAI)-3`.CNJ

o-mis-a
3SG.POSS-elder.sister(NDA)-OBV
'In this way he immediately knows himself that his older sister is a killer.' (BR)

Given this frequency of interdialectal exchange, speakers and linguists alike report a high degree of mutual intelligibility between Western Cree dialects. For example, L1 Woods Cree speaker and decorated educator Solomon Ratt said the following in his 2016 (Plains Cree) textbook *mâci-nêhiyawêwin*:

> "In Saskatchewan, speakers of the three dialects found here, the N [Swampy], Y [Plains] and TH [Woods], can communicate in Cree, but there are certain differences in word usages that sometimes need explanation. No dialect is better than any other." (Ratt 2016:1)

Prominent Cree linguist Arok Wolvengrey mirrors this sentiment, likening the difference between Western Cree dialects to those between dialects of English:

> "In addition to sound variation, each of the dialects has a certain amount of difference in vocabulary, just as is found between British, Canadian and Australian English. However, all three Cree dialects of Saskatchewan represent a single language which speakers can, with varying degrees of difficulty, readily understand." (Wolvengrey 1998:117)

For further accounts of the mutual intelligibility of Western Cree, see Appendix B, in which I have compiled excerpts from interviews with L1 speakers of various Western dialects, in which these speakers discuss their experience with cross-dialectal communication. These interviews largely corroborate the quotations provided above; barring some difficulty in immediate acclimatisation, communication between speakers of differing dialects is non-problematic.

Morphosyntactically speaking, the Western Cree group may be divided into conservative dialects, which have preserved the great majority of inflectional paradigms from Proto-Algonquian, and innovative dialects, which have either lost many of these paradigms or syncretically collapsed them together. Swampy Cree and (especially) Moose Cree are generally considered to be conservative dialects (Kang 2017:12; Oxford 2017:4), while Plains Cree is seen as paradigmatically innovative or "simplified" (Bakker 2006:4). Broadly, Woods Cree also falls into the morphosyntactically innovative category, having lost many of the paradigms still preserved in its eastern neighbours. However, particular in its eastern varieties, Woods Cree tends not to be as morphosyntactically innovative as Plains Cree, and still preserves, even if only

vestigially, some inflectional contrasts now lost in Plains Cree. As such, although Woods Cree can generally be considered an innovative variety, it serves as something of a middle point between the highly innovative Plains Cree and the more strictly conservative Swampy Cree and Moose Cree. In this way, the morphosyntactic innovation (or simplification) of the Western Cree dialect chain can be seen to broadly increase as one moves westward:

More Innovative	<<<	<<<	<<<	<<<	*More Conservative*
	Plains Cree	Woods Cree	Swampy Cree	Moose Cree	
West	>>>	>>>	>>>	>>>	*East*

Figure 4. A rough approximation of patterns in morphosyntactic innovation in Western Cree dialects

As a final means of illustrating the similarities and differences between Woods Cree (WC) and its neighbours (PC and SC), I have fabricated a short, tridialectal dialogue using attested L1 speaker utterances found in the Algonquian Linguistic Atlas (Torkornoo et al. 2021). This dialogue may be found in full in Appendix C; for demonstrations purposes, I will also provide a brief excerpt below:

PC. *môya, apisîs piko nitayâwâw sôniyâw. kiya mâka, kîkwây ê-nitawêyihtaman?*

môya	apisîs	piko	nit-ayâw-âw	sôniyâw	kiya	mâka
NEG	little	only	1SG>3SG.IND-have(VTA)-1SG>3SG.IND	money(NA)	2SG	but

kîkwây	ê-nitawêyiht-aman
what	CNJ-want(VTI)-2SG.CNJ

WC. *namwâc, apisîs poko nitayâwâw sôniyâw. kîtha mâka, kîkwây kâ-nitawîthihtaman?*

namwâc	apisîs	poko	nit-ayâw-âw	sôniyâw	kîtha	mâka
NEG	little	only	1SG>3SG.IND-have(VTA)-1SG>3SG.IND	money(NA)	2SG	but

kîkwây	kâ-nitawîthiht-aman
what	CNJ-want(VTI)-2SG.CNJ

SC. *môna, apišîš piko nitayâwâw šôniyân. kîna mâka wîna, kêkwân wâ otinikêyan?*

môna	apišîš	piko	nit-ayâw-âw	šôniyân	kîna	mâka
NEG	little	only	1SG>3SG.IND-have(VTA)-1SG>3SG.IND	money(NA)	2SG	but

wîna	kêkwân	wâ	otinikê-yan
FOC	what	DUB	get(VAI)-2SG.CNJ

‘No, I only have a little money. But what about you, what are you getting?’

As demonstrated by the numerous accounts above, although certainly not identical, the Western Cree dialects closely resemble one another in both structure and lexicon, possessing a high degree of mutual intelligibility.

1.4 Dialectal Variation within Woods Cree

In addition to variation between Woods Cree and its neighbours, there is also considerable variation within Woods Cree, primarily concerning morphophonology and the lexicon. As such, throughout this investigation, I will split Woods Cree into two broad varieties, Eastern Woods Cree and Western Woods Cree[4]. Correspondingly, I will draw my language data from a corpus consisting of materials from a balanced group of both Eastern and Western communities (see Section 1.6). For Western Woods Cree, I will use language data from the towns of La Ronge (LR, *mistahi-sâkahikanihk*, pop. 2688[5]), and Stanley Mission (SM, *âmaciwîspimowinihk*, pop. 104). For Eastern Woods Cree, I will use data from the towns of Pukatawagan (PK, *pakitahwâkanihk* pop. 1724) and South Indian Lake (SL, *opiponapiwin*, pop. 981). In addition to these paradigmatically Western and Eastern communities, I will also incorporate a smaller amount of language data from communities near the geographic midpoint of the Woods Cree dialect zone, demonstrating both Eastern and Western characteristics. As representatives of these 'Intermediate' varieties, I will use data from the towns of Pelican Narrows (PN, *opâwikoscikanihk*, (pop. 2703), Brochet (BR, *kîsipakamâhk*, pop. 2410), and Kinoosao (KN, *kinosîw*, pop. 58).

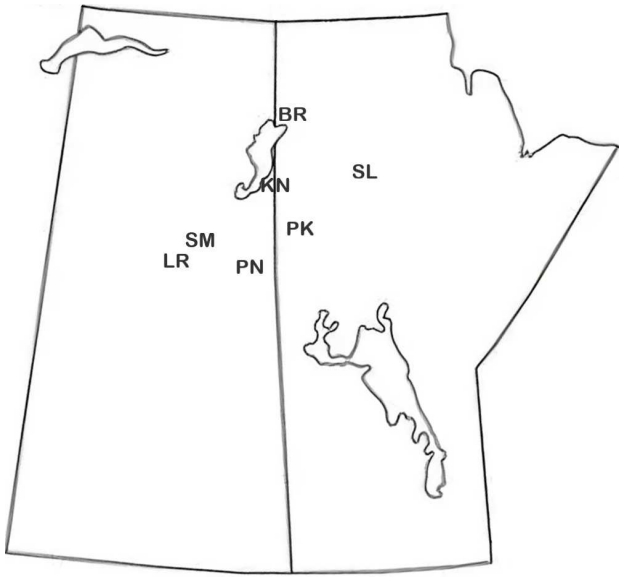

Figure 5. Communities from which the majority of the language data in this investigation stems (LR - La Ronge, SM - Stanley Mission, PN - Pelican Narrows, KN - Kinoosao, BR - Brochet, PK - Pukatawagan, SL - South Indian Lake)

[4] Pentland (1978) discusses a similar partition, albeit in less detail.
[5] All population data is taken from the 2021 Canadian census.

Broadly speaking, Eastern Woods Cree tends to be more morphosyntactically conservative (similar to Swampy Cree), while Western Woods Cree tends to be more innovative (similar to Plains Cree). Below, I will outline below a selection of conspicuous East-West isoglosses, and therewith establish a basic intradialectal map for Woods Cree. Labelled isogloss maps for each feature discussed may be found in Appendix D.

mihko - misko

One notable lexical variance between Western and Eastern varieties of Woods Cree is their pronunciation of the word for 'blood' (from PrA *meçkwi). Among Western speakers, both *mithko* and *misko* are common usage (although *misko* predominates among younger speakers), whereas in Eastern Woods Cree, *mihko* is the dominant form. The isogloss for this feature appears to occur roughly at the political boundary between Saskatchewan and Manitoba; in the intermediate community of Pelican Narrows, for example, both *misko* and *mihko* are attested, while in Brochet, only *mihko* is attested. A similar pattern may been seen in the contemporary realisations of several other lexemes containing reflexes of the Proto-Algonquian *ç (or *x); this is discussed in more detail in Section 2.2.1:

3) Proto-Algonquian *mo·çkaho·siwa ('crane')
   Plains Cree: *môhkahosiw*
   Woods Cree:
      La Ronge: *môthkahosiw*
      Stanley Mission: *môthkahosiw*
      Pukatawagan: *môhkahosiw*[6]
   Swampy Cree: *môhkahosiw*

ikwa - akwa

ikwa is an extremely frequent lexical and grammatical particle with a wide variety of functions; broadly, however, it may be understood as a coordinating conjunction, similar to the English 'and'. The form *ikwa* exists in both Western and Eastern Woods Cree; however, while *ikwa* is the only form attested in Western Woods Cree, Eastern Woods Cree also has the variant *akwa*, which occurs in free variation with *ikwa*. Both forms are attested in South Indian Lake, Pukatawagan, and Brochet, while only *ikwa* is attested in La Ronge, Stanley Mission, and Pelican Narrows.

nâ - cî

There exist three 'yes-no' interrogative particles in Woods Cree; *nâ*, *cî*, and *ciyi*. *nâ* and *cî* appear largely in free variation in La Ronge and Stanley Mission (with perhaps a slight preference for *cî*), however in South Indian Lake and Pukatawagan, *nâ* is much more frequent, with *cî* only being used in questions 'in which the speaker expects an answer which confirms his/her

[6] In practice, this is most often seen in the term *misi-môhkahosiw* (lit. 'large crane')

expectations' (Starks 1992:161), if at all. Although both *nâ* and *cî* are found in neighbouring Cree dialects, Woods Cree is unique in having both forms, and has had both since at least the early 19th century (Howse 1844: 279). By contrast, in neighbouring Plains Cree, only *cî* is used, whereas in Swampy Cree, only *nâ* is used.

***kâ-* Reduction**

The grammatical prefix *kâ-* is one of several prefixal markers for Conjunct Order verbs (see Section 2.3.4.2.2), and is of extremely frequent use in both Eastern and Western Woods Cree. However, in Pukatawagan and South Indian Lake, the initial /k/ is optional, and the prefix may be realised simply as *â-*. In La Ronge, Stanley Mission, Brochet, Pelican Narrows, and Kinoosao, this elision is entirely unattested.

kîkway - kîkwan

The interrogative pronoun *kîkway* (see Section 2.3.2.3), is attested across Woods Cree; however, in Pukatawagan, South Indian Lake, and Kinoosao, *kîkway* is in free variation with *kîkwan*, a pronoun of identical function.

***na-* Future Tense Marker**

The first-person future tense marker *na-* (see Section 2.3.4.3), is a peculiarity of Eastern Woods Cree, and is entirely absent in the Western variety, which uses *nika-* instead. La Ronge, Stanley Mission, Pelican Narrows and Brochet both use the Western form *nika-*, whereas Pukatawagan, South Indian Lake, and Kinoosao use the Eastern *na-*.

	La Ronge	Stanley Mission	Pelican Narrows	Brochet	Kinoosao	Pukatawagan	South Indian Lake
mihko ~ misko (mithko)	*misko mithko*	*misko mithko*	*mihko misko*	*mihko*	*?*	*mihko*	*mihko*
ikwa ~ akwa	*ikwa*	*ikwa*	*ikwa*	*ikwa akwa*	*?*	*ikwa akwa*	*ikwa akwa*
nâ~cî	*cî>nâ*	*cî>nâ*	*?*	*?*	*?*	*nâ>cî*	*nâ>cî*
kâ- ~ (k)â-	*kâ-*	*kâ-*	*kâ-*	*kâ-*	*kâ-*	*(k)â-*	*(k)â-*
kîkway ~ kîkwan	*kîkway*	*kîkway*	*kîkway*	*kîkway*	*kîkway, kîkwan*	*kîkway, kîkwan*	*kîkway, kîkwan*
nika- ~ na-	*nika-*	*nika-*	*nika-*	*nika-*	*na-*	*na-*	*na-*

Table 3. Isoglossic features and boundaries (indicated in bold) between Western Woods Cree (on the left) and Eastern Woods Cree (on the right).

Tabularising these isoglosses, a clear pattern in East-West variation is seen; La Ronge and Stanley Mission consistently pattern together in opposition to Pukatawagan and South Indian Lake, while Pelican Narrows, Brochet, and Kinoosao all exhibit traits of both varieties to varying degrees. Based on this, it may be concluded that the overall boundary between Eastern and Western Woods Cree is situated somewhere near the provincial border of Saskatchewan and Manitoba, with the intermediate region between the two containing much of the aforementioned Peter Ballantyne Cree Nation.

1.5 Existing Woods Cree Literature

In comparison with its more widely spoken neighbours, Woods Cree has seen comparatively little linguistic documentation in both academic and lay circles. In her 1992 doctoral thesis, *Aspects of Woods Cree Syntax*, one of few modern descriptive analyses of the language, Donna Starks refers to Woods Cree as "the least studied" of the Western Cree dialects (Starks 1992:8), with other authors noting Woods Cree as "yet to be studied in detail" (Greensmith 1985:2). Although, as will be discussed below, documentary studies of Woods Cree do exist, unlike its neighbours Plains Cree (see Wolfart 1973) and Swampy Cree (see Ellis 2000; 2004;2016), there remains to the present day no widely-accepted 'definitive' descriptive grammar of Woods Cree.

Aside from a small number of references in historical texts, the first major documentation of Woods Cree came in 1844 with the publication of the *Grammar of the Cree Language, with Which is Combined an Analysis of the Chippeway Dialect* by Joseph Howse. As implied by the name, this 'ground-breaking grammar' (Pentland 2003:287) provides parallel morphosyntactic descriptions of both (Woods) Cree and 'Chippeway' (now known as Ojibwe), albeit with considerably more focus towards the former. Although still a useful point of entry for documentation, the use of the Howse grammar as a reference for modern Woods Cree should be approached with reticence. Aside from being written 180 years ago, the variety of Woods Cree described in the grammar, a subdialect spoken in Eastern Manitoba, no longer exists, having been supplanted by Swampy Cree. For these reasons, I will primarily be consulting the Howse grammar as a reference for historical Woods Cree, rather than as a primary source for the language today.

Aside from the Howse grammar, the closest thing to a comprehensive, modern descriptive grammar of Woods Cree is the aforementioned 1992 doctoral thesis of Donna Starks. In addition to detailing the language's syntax, Starks also dedicates a substantial portion of this thesis towards general documentation of phonology, orthography, and morphology. However, *Aspects of Woods Cree Syntax* is not intended to be a generalised description of Woods Cree as a whole, instead limiting its descriptive efforts specifically to the dialect of Woods Cree spoken at South Indian Lake. In Stark's own words, "no attempt should be made to generalize this description to the entire Woods Cree dialect area." (Starks 1992:10). Despite this, *Aspects of Woods Cree*

Syntax still represents the most extensive modern grammatical description of Woods Cree to date, and will be referenced extensively throughout this text.

Two further major works in the documentation of Woods Cree are the *Memoirs of the Elders of Pukatawagan* and the accompanying *Westfall & Castel English-Cree Dictionary*, both written in 2001 by Robert Castel and David Westfall. The former of these is a large collection of personal reminiscences from various Woods Cree elders, which I will discuss in further detail in Section 1.6.1.3. The latter, the *English-Cree Dictionary*, derives its contents entirely from this collection, being an alphabetised list of all wordforms found in the *Memoirs*, each provided with an English definition and part-of-speech tag, with all entries written in the Standard Roman Orthography (see Section 2.1.1). In total, this dictionary contains ~12600 entries; however, in addition to these entries, it also contains a brief grammatical description of Woods Cree in its foreword and appendices, as well as morpheme lists, inflectional paradigms, and other metalinguistic information. Although the contents of the *Memoirs of the Elders* are available online, the *English-Cree Dictionary* is not, and is presently accessible only in print.

The only other major dictionary of Woods Cree is the *Colin Charles Cree Dictionary* (or CCD), a bilingual dictionary originally written in the 1970s by fluent speakers Colin Charles and Keith Goulet of the Lac La Ronge Indian Band, and later revised, updated, and digitised during the 2010s by Minnie McKenzie, another fluent speaker of the same band. The dictionary is split into two sections, Cree-English and English-Cree, of which I have primarily made use of the Cree-English section, which contains 5858 entries. Owing to the provenance of its authors, the CCD primarily records the variety of Woods Cree spoken at La Ronge, however, it also includes words from various neighbouring communities, with entries written in the Colins Modified Orthography (see Section 2.1.2). In addition to its lexical entries, the CCD also contains a small number of inflectional paradigms and example sentences, and some cursory grammatical description. The contents of the more recent updated edition have been digitised, but are not publicly available online. Instead, I gained access to them through a signed Memorandum of Understanding between the Lac La Ronge Indian Band (who own the rights to the dictionary), the Alberta Language Technology Lab (ALTLab), and the First Nations University of Canada, wherein the parties agreed to "collaborate in the collection and creation of various linguistic materials, including the development of language technological software tools and applications for *nîhithawîwin*".

Various educational materials concerning Western Woods Cree morphology have also been created by Ben Godden, a Saskatchewan-based linguist and L2 Woods Cree speaker; in this thesis, I will chiefly be making use of an extensive set of paradigm tables published online by Godden in 2019, based on the subdialect of Stanley Mission[7].

[7]https://creeliteracy.org/2019/06/18/summary-paradigm-tables-the-decoder-ring-of-cree-verbs-y-and-th-dialects/

One of the only other major academic documentations of Woods Cree is Jennifer Greensmith's 1985 thesis *Phonological Variants in Pukatawagan Woods Cree*. As its name implies, this text is predominantly concerned with phonological processes and the underlying phonemic inventory of (Eastern) Woods Cree; however, it does contain some morphological information, including relevant morphophonological processes.

Finally, although not documentations of Woods Cree, I will include here reference to H. C. Wolfart's 1973 treatise *Plains Cree: A Grammatical Study*, as well as the textbook and descriptive grammar *Spoken Cree* by C. Douglas Ellis, which describes Swampy Cree and Moose Cree. These works are widely viewed as the definitive descriptive grammars of their respective dialects, and I will refer to them frequently throughout Section 2 to illustrate similarities and contrasts between Woods Cree and its neighbours. These descriptive sources on Woods Cree and its neighbours will serve as the basis for most of my claims concerning morphosyntax and phonology in Section 2, where I will verify and elaborate on the observations in these works using Woods Cree corpus data (see Section 1.6). Any phenomena which cannot be adequately documented using pre-existing descriptive sources or the corpus will instead be described based on discussion with an L1 Woods Cree consultant (Solomon Ratt, a male speaker in his early 70s from Stanley Mission).

1.5.1 Other Documentation

Outside of academia, few Woods Cree language-learning resources exist for public consumption. The two largest online learning and vocabulary resources for Woods Cree are https://learncree.ca, a self-teaching resource hosted by the Lac La Ronge Indian Band, N-12 Education Program, and the https://creeliteracy.org, a blog-style Cree language advocacy website. The former of these contains 22 illustrated vocabulary lessons on a variety of semantic domains, each containing ~10-30 distinct lexical items with accompanying audio. In addition, the site has audio and lyrics for 17 educational songs, and a catalogue of several dozen educational books and posters available for physical purchase. The latter is primarily a collection of Cree stories and blog posts (see Section 1.6.1.1); however it does also contain a number of posts intended for language learners, containing Cree vocabulary with English translations and descriptions of grammatical phenomena. However, virtually all of these posts provide descriptions of Plains Cree, rather than Woods Cree; only 3 such educational posts, with a total of 99 example words, are solely in Woods Cree.

Finally, a number of Woods Cree educational apps intended for mobile devices exist, in varying states of continuing support. The largest and most accessible such app is the Woodland Cree Conversation App, an app developed by Delasie Torkornoo, Marie-Odile Junker, and Claire Owen of the Algonquian Dictionaries and Language Resources Project

(https://apps.apple.com/ca/app/woodland-cree-conversation/id1581282722)[8]. Available on both Apple and Android devices, this app contains a several hundred phrases in Woods Cree, rendered in Standard Roman Orthography (see Section 2.1.1) with accompanying audio from at least two L1 speakers (Minnie McKenzie, Martha Michell, and Hilda Morin) from Woods Cree communities across Saskatchewan (Stanley Mission, Kinoosao, and Southend respectively), with these phrases being given translations in both English and French. In addition to this app, there exist two similar apps developed by the Little Red River Cree Nation and Woods Cree settlement at Deschambault Lake, both purportedly containing over 1000 vocabulary items each (https://www.pagc.sk.ca/dl-woodland-cree-app). These apps, however, are no longer usable on modern devices, having obsolesced due to a lack of support from the developers.

As these examples illustrate, although some online language resources for Woods Cree do exist, they are limited in size and scope, and are scattered across various platforms (some of which are no longer available). As such, much of the remainder of this paper will discuss remedying this lack of resources.

1.5.2 Existing Online Language Resources in Other Western Cree Dialects

In contrast to Woods Cree, a (relatively) large number of online language tools exist for the other Western Cree dialects, particularly Plains Cree and Swampy Cree. Compared to Woods Cree's zero, there exist three major Swampy Cree dictionaries online[9], as well as several language learning apps and websites. Plains Cree has even further resources, with two major online dictionaries containing tens of thousands of entries between them[10], including an intelligent dictionary with in-built morphological analyser and paradigm generator, as well multiple self-guided language-learning websites and apps. Thus, it is clear that the reason behind the lack of comparable resources for Woods Cree is not due to technological limitations, but simply to a lack of allocated resources. These tools are to be discussed in much further detail in Section 3.

1.6 Corpus

As illustrated at length in Section 1.5, although the literary documentation of Woods Cree is sparse, the language is not entirely without pre-existing written resources. However, to date, there exists no centralised database of these resources, and therefore no searchable corpus with which to efficiently verify the contents of these works. As such, for the purposes of this investigation, it was necessary to construct such a corpus as a point of reference.

---

[8] Alternatively, one may find the contents transcribed here: https://www.eastcree.org/pdf/CONVERSATION_Woodland.pdf

[9] https://www.spokencree.org/glossary
https://dictionary.moosecree.atlas-ling.ca/
https://fortsevern.atlas-ling.ca/

[10] https://www.creedictionary.com/
https://itwewina.altlab.app/

The resultant corpus, which I will dub the Miscellaneous Woods Cree Text Corpus (MWCT, or in this thesis, simply 'the text corpus'), consists of four major textual sources; the Cree Literacy Network (CLN), the Government of Canada (GC), the Memoirs of the Elders of Pukatawagan (ME), as well as a number of smaller texts (MISC), all of which also contain full English translations. Each of these sources is discussed in detail in Section 1.6.1. The methods involved in the construction and formatting of this corpus, as well as the nature of its contents, were heavily inspired by those used to create the Miscellaneous Cree Texts Corpus (MCT), the largest extant corpus of Plains Cree, discussed in detail in Dacanay & Arppe (forthcoming). Eventually, I plan to merge this Woods Cree corpus, and a similar, in-progress text corpus of Swampy Cree, with the MCT in order to create a broader corpus of Western Cree in general; indeed, I will occasionally be drawing examples from these Plains and Swampy Cree corpora throughout this investigation for the sake of comparison. However, as might be expected, I will be drawing most heavily from the Woods Cree corpus throughout this thesis; unless otherwise specified, all future example words and sentences originate either from this corpus or from the CCD. A further outline of the formatting conventions which I will use when drawing examples from the corpus is provided in the preface on page xiv.

1.6.1 Corpus Sources

1.6.1.1 Cree Literacy Network (CLN)

The Cree Literacy Network (or CLN) is a language advocacy group, founded with the goal of promoting language use and literacy across Cree dialects, as well as being a major proponent of the usage of the Standard Roman Orthography (see Section 2.1.1). The main website of the Cree Literacy Network, https://creeliteracy.org, is one of the largest online repositories of Cree text, containing transcriptions of stories, songs, prayers, educational materials, and blog-style posts in a variety of Cree dialects. Most of the Woods Cree contents of the CLN website are written by Solomon Ratt, an L1 speaker from Stanley Mission, Saskatchewan. As mentioned, the materials produced by the CLN are already publicly available through their website; however, I have also received explicit permission from Arden Ogg, the director of the organisation, to make use of these texts for the purposes of corpus creation. In total, the 172 Woods Cree texts from the CLN constitute 47197 tokens, representing 23.4% of the total.

1.6.1.2 Government of Canada (GC)

Since the late 2000s, the federal government of Canada has been, at irregular intervals, publishing Indigenous language translations of various informational documents, governmental declarations, and other textual miscellanea. Woods Cree document translations represent a small minority of this set, with these texts being scattered across the federal government's various official websites with no centralised database or consistent tagging scheme. As such, these texts were collected by using various high-frequency Woods Cree lexemes as site-wide search queries on the various governmental websites containing such translations; for a more detailed

description of this manual text collection method, see Dacanay & Arppe (forthcoming). In total, I collected 11 Woods Cree texts from various governmental websites, totalling 17550 tokens, which represents 8.7% of the total. These governmental texts all appear to have been written by speaker(s) of the Western variety of Woods Cree, however, the name(s) and place(s) of origin of the author(s) are not known for any of the texts in this subcorpus.

1.6.1.3 Memoirs of the Elders of Pukatawagan (ME)

As their name implies, the *Memoirs of the Elders of Pukatawagan* (ME) are a collection of autobiographical tales, reminiscences, and anecdotes told by various elders from Pukatawagan, Manitoba, including several monolingual or quasi-monolingual Woods Cree speakers. These stories were elicited through interviews conducted entirely in Woods Cree by various residents of Pukatawagan between 1999 and 2001, these interviews being compiled, edited, and transcribed by the linguist David Westfall, with translations by Robert Castel, a Woods Cree native speaker. As mentioned, ME was written as a companion text to the *Westfall & Castel English-Cree Dictionary*, with this dictionary's contents ultimately stemming from the contents of ME. In addition, five years after ME was published, a shorter, supplementary text, titled *Speaking to the Future: Additional Memoirs of the Elders of Pukatawagan*, was published, containing an additional 5 interviews and an accompanying glossary; however, while I have obtained a physical copy of this text, I have yet to locate a digitised version. As such, I have excluded its contents from the corpus for the time being. I was unable to contact David Westfall directly concerning explicit usage permissions for the corpus on account of his having deceased several years prior to the composition of this thesis; however, the copy of ME which I used for the corpus was freely available in full on what appears to have been his personal website[11].

Over the course of the main text of ME, a total of 18 elders were interviewed, ranging in age from 53 to 87. These interviews covered a variety of topics, mostly concerning how life in Pukatawagan had changed over the course of the 20th century, and the contrast between Cree life before and after contact with Europeans; however, all 18 interviews were at least semi-autobiographical in nature. For each speaker, I compiled all of their interviews into a single text, resulting in a total of 18 texts containing 118844 tokens (59% of the total) between them.

1.6.1.4 Miscellaneous (MISC)

In addition to these three major sources, several other, smaller written works of Woods Cree were also included in the corpus. These include example sentences and transcribed conversations from the works of Donna Starks (1992), representing speech from South Indian Lake, the transcribed stories and example sentences of Robert Brightman (1985), an anthropologist who studied various Woods Cree communities, most notably Brochet, the transcribed Woods Cree stories included in Ida Swan's 2001 thesis, representing speech from Pelican Narrows, and the short anthology 'Woods Cree Stories', a collection of humorous anecdotes from Stanley Mission,

[11] https://hillmanweb.com/westfall/

written and compiled by Solomon Ratt. In addition to these, various pieces of Woods Cree literary ephemera, such as captions and flavour text from websites, community safety pamphlets, and public blog posts, were also included, all of which were located by simply entering high-frequency Woods Cree words as Google search queries (see Dacanay & Arppe, forthcoming). In total, this miscellaneous subcorpus contained 32 texts, comprising a total of 17729 tokens, representing 8.8% of the total.

1.6.2 Internal Format

Initially, each constituent text in the corpus was converted into a UTF-8 encoded .txt file. This file contained the source URL of the relevant text followed by the full contents of the text in Cree and English on alternating lines, with each line typically representing a single sentence. These bilingual sentence pairs were separated from one another with an empty line. In instances in which a given English sentence did not have an exact corresponding sentence in Cree, or vice-versa, multiple sentences were given on a single line until a coterminous, discursively equivalent unit could be found in the text in the other language.

```
>https://creeliteracy.org/2022/10/13/think-nothing-of-it-so
lomon-ratt-th-dialect/

That’s it then, think nothing of it:
ikosi mâka, kâwitha nânitaw itîthihta:
```

From this manually compiled, human-readable format, these files were further computationally processed into the VRT format (Evert & Hardie 2011), allowing for certain pieces of primary metadata (such as the source URL and subcorpus) to be stored directly in the text using XML markup. In this format, the manually aligned English and Cree sentences were represented as paragraphs (`<p id=...>`), with each line overtly marked with the ISO 639 code for the relevant language (`cwd` for Woods Cree and `eng` for English). Each Cree line was further partitioned into individual sentences, with each sentence on its own line in a section labelled (`<s id=...>`). These VRT files were subsequently morphologically analysed using the Woods Cree morphological model discussed in Section 3.4.1, with each tokenised Cree word separately provided with a morphological analysis on its respective line[12].

[12] These analyses were further scrutinised using a Constraint-Grammar-based disambiguator designed for Plains Cree (Schmirler et al. 2018; Arppe et al. 2020), which selected the most syntactically plausible analysis for each lexeme based on surrounding lexical context. Developing a similar disambiguator specifically for Woods Cree is an avenue for future research.

```
<text id="CLN-131022">
<url="https://creeliteracy.org/2022/10/13/think-nothing-of-
it-solomon-ratt-th-dialect/">
<p id=1>
<crk>ikosi mâka, kâwitha nânitaw itîthihta:</crk>
<eng>That’s it then, think nothing of it:</eng>
<s id=1>
ikosi     îkosi     _+Ipc          îkosi
mâka mâka _+Ipc          mâka
,    ,    _+CLB          ,
kâwitha   wîhthîw
PV/ka+_+V+TA+Cnj+12Pl+3SgO+Var/East+Err/Orth @PRED-TA
[ka-wîhthahk]
nânitaw   nânitaw   _+Ipc          nânitaw
itîthihta itîthihtam     _+V+TI+Imp+Imm+2Sg  @PRED-TI
itîthihta
:    :    _+CLB          :
</s>
</p>
```

1.6.3 Metadata

As mentioned, three pieces of metadata are stored internally in each VRT file, these being the filename, the source URL, and the subcorpus. Filenames within the corpus are composed according to the formula SC-DDMMYY, with SC being the subcorpus and DDMMYY representing the date of publication. For example, a file in the Miscellaneous subcorpus written on the 4th of July, 2019 would be represented with the filename MISC-040719. Texts for which the exact date of composition is unknown are named according to the formula SC-n, whereby n is an integer increasing by one for each such text (e.g. MISC-1, MISC-2, etc.). Aside from this, all additional metadata on each constituent file is stored in an external metadata sheet. With some (intentional) redundancy, this sheet contains the filename, source URL, date of creation, genre, author (including date and place of birth), SRO rating[13], subcorpus, and original language of each file. Complete metadata could be found for 205 of the 233 total texts within the corpus. The 28 texts with incomplete or absent metadata account for 13.3% of the total tokens in the corpus. Bearing this limitation in mind, I will draw examples from these texts in my analysis only when examples cannot be found elsewhere, and will explicitly note any such cases in the interlinear gloss with the geographic tag UL (for Unknown Location).

[13] A number between 1 and 4 representing how closely the orthography resembles the Standard Roman Orthography (see Sections 2.1.1 and 2.2.1), with 1 being wholly idiosyncratic and 4 being wholly standard

1.6.4 Statistics

In total, the MWCT contains 201320 tokens across its 233 texts, with 86224 tokens of Woods Cree text and 115096 tokens of English text. Removing punctuation and other non-lexical content, the total number of Cree word tokens shrinks to 63968, representing 18650 distinct word types.

Community	Number of Texts	Total Token Count (%)
Pukatawagan	18	118844 (59%)
Stanley Mission	170	51794 (23.4%)
South Indian Lake	11	5243 (2.6%)
Pelican Narrows	9	4553 (2.3%)
Brochet	3	2913 (1.4%)
Montreal Lake[14]	2	586 (0.3%)
Southend	3	545 (0.3%)
La Ronge	1	499 (0.2%)
Norway House[15]	2	382 (0.2%)
Unknown Location	14	17829 (8.9%)

Table 4. Number of texts and associated token counts from different communities in the MWCT corpus

In terms of genre, the corpus is dominated by personal reminiscences, which I define as any autobiographical anecdote or personal history, which constitute 47 out of 233 texts and 129142 tokens. After this, poems, which I loosely define as intentionally metrical writing created for the purpose of artistic and aesthetic expression, constitute a large number of individual texts (47/233), but ultimately a relatively small total token count (3388). Blog posts, which I define broadly as any informal, natively-online text intended for popular consumption, constitute 28 out of 233 texts and 8845 tokens, followed by stories, which constitute 25 out of 233 texts and 10423 tokens. The exact genre boundary between stories and personal reminiscences was somewhat permeable; generally, I classed a text as a ‘story’ if it contained little or no autobiographical information, or if it was otherwise a work of fiction. Legends, by contrast, were defined as any story containing clear mythological or otherwise fantastical elements, as well as any traditional stories not confabulated by the speaker. These legends made up 20 out of 233 texts, constituting

[14] A set of two reserves in central Saskatchewan with a combined population of 1555

[15] A primarily Swampy Cree town of roughly 5000 inhabitants, located near the Woods-Swampy dialect boundary on the northern shores of Lake Winnipeg, Manitoba

20972 tokens. The remainder of genres, mostly informational material such as pamphlets, phrase translations, prayers, and song lyrics, constitute 68 out of 233 texts and 24023 tokens.

Genre	Number of Texts	Token Count (Types)
Personal Reminiscence	47	129142 (64.1%)
Poem	47	3388 (1.7%)
Blog Post	28	8845 (4.4%)
Story	25	10423 (5.1%)
Phrase Translation	23	1286 (0.6%)
Legend	20	20972 (10.4%)
Informational Pamphlet	10	14498 (7.2%)
Prayer	8	631 (0.3%)
Song	8	1500 (0.7%)
Public Service Announcement	6	3711 (1.8%)
Interview	5	3241 (1.6%)
Miscellaneous[16]	6	442 (0.2%)

Table 5. Number of texts and associated token counts of various genres in the MWCT corpus

The vast majority of texts followed the orthographic conventions of the SRO; of 233 texts, 227 were given an SRO rank of 4 (the highest possible), with the remaining 6 given a rank of 3. This is not, however, to say that the 227 texts with ranks of 4 were orthographically identical; they differed in their representation of the phoneme /ð/ (by either ⟨th⟩ or ⟨ð⟩), their means of marking vowel length (by either macron or circumflex (see Section 2.1.1)), and their inclusion (or exclusion) of short unstressed vowels (see Section 2.2.3), among other minor differences. This orthographic diversity was retained in the corpus for reasons of descriptive faithfulness, although standardised versions of each lexeme exist in the morphological tagging (see Section 1.6.2).

1.6.5 Expansion

Although I am confident in stating that the current version of the MWCT contains the majority of written Woods Cree currently available on the surface web, it is by no means an exhaustive

[16] This miscellaneous category includes 2 Woods Cree forewords from academic papers, 2 transcripts of unscripted conversations, 1 personal correspondence letter, and 1 dedication plaque from a memorial

repository of all extant Woods Cree text. In future, I plan to expand this corpus by digitising the contents of *Speaking to the Future* and the example sentences used in the *Westfall & Castel English-Cree Dictionary*, as well as by isolating and extracting example sentences from the *Colin Charles Cree Dictionary*. In the long term, the Woods Cree contents of the Howse grammar, with some degree of orthographic modernisation, would also serve as an obvious source for corpus expansion, as would negotiating with the creators of the Woodland Cree Conversation App to discuss incorporating its contents into the corpus. Furthermore, the Cree Literacy Network and Government of Canada continue to publish texts in Woods Cree with varying regularity; thus as new texts arise, the corpus is ripe for continual updating.

1.6.6 Accessibility

The MWCT has been made available for non-commercial research use through the Alberta Language Technology Lab (ALTLab), and may be accessed upon request by contacting my email (dacanay@ualberta.ca). In future, it will also be made publicly available through a Git repository.

1.6.7 Audio Corpus

In addition to the text corpus described above, I have also compiled an audio corpus of Woods Cree, consisting entirely of publicly available spoken audio already present on the internet. This audio corpus is much more provisional than the text corpus, and has not yet been transcribed, translated, or annotated. Instead, it consists of a spreadsheet directory containing 104 audio recordings of fluent Woods Cree with accompanying metadata (namely, original source, name(s) of speaker(s), place(s) of origin of speaker(s), genre, date of recording, and length). Most of the audio recordings contained in the corpus are embedded in videos, some of which have English subtitles; however, as mentioned, there remains no external (Cree) transcription of its contents. As such, it is impossible to ascertain exactly how large this audio corpus is in terms of token count. The total length of the audio contained in the corpus is ~21 hours, however, this includes some English dialogue, pauses, and other non-speech events.

The contents of the audio corpus primarily consist of interviews with various L1 Woods Cree-speaking elders, conducted throughout the late 20th and early 21st centuries. A substantial portion of the audio stems from several collections of interviews taken by Native Communications Inc., a public radio broadcaster in Manitoba. These interviews consist of elders from various communities across western Canada providing autobiographical anecdotes, as well as speaking of how the Cree lifestyle has changed since their youth, similar to the *Memoirs of the Elders* in the text corpus. In addition to these interviews, a large number of the entries in the corpus are recitations of traditional stories and legends. The remainder of the audio consists largely of translated public service announcements and educational videos (such as songs). Like the materials in the text corpus, the totality of the audio corpus stems from existing freely

available online materials, with my chief contribution therefore being the location, centralisation, and provision of consistently formatted metadata for these materials.

As mentioned, the audio corpus is highly provisional in comparison to its textual counterpart and remains a work-in-progress, with several major sources of (freely accessible) audio yet to be included. One of these sources is *Achimowin*, a Woods Cree radio news program which has been broadcast several times each week by Missinipi Broadcasting Corporation (or MBC) since 1984. The archives of the MBC are available online[17], containing audio from nearly 6000 distinct broadcasts; although only a fraction of these are broadcasts of *Achimowin*, these archives nonetheless likely contain hundreds of hours of Woods Cree audio, including guest speakers from across the Woods Cree dialect area. The second major source of Woods Cree audio is the Cree Literacy Network. Many of the Woods Cree written materials described in Section 1.6.1.1 have accompanying audio; with some parallelisation, these paired written and spoken materials could form the basis of a future, multimodal Woods Cree corpus. However, these recordings, and the MBC materials, have been excluded from the audio corpus in this investigation, purely on the basis of time constraints. Owing to its incompleteness, the audio corpus is not yet publicly available; however, once a Git repository is established for the text corpus, the metadata sheet for the audio corpus will accompany it. Despite this incompleteness, the audio corpus will nonetheless serve as the primary data source for phonological information throughout this investigation; unless otherwise specified, pronunciations and IPA transcriptions provided throughout are sourced from audio found in this corpus.

[17] https://omny.fm/shows/mbc-extra

2. A GRAMMATICAL DESCRIPTION OF WOODS CREE

In this section, I will provide a descriptive overview of various linguistic features of Woods Cree, noting therein subdialectal variation and providing comparisons throughout with neighbouring dialects and historical varieties. I begin by describing Woods Cree orthography, before moving to a discussion of the language's phonology, including its phoneme inventory, common phonemic processes, syllable structure, and stress. In this subsection, I will draw my data chiefly from Greensmith (1985a) and from the aforementioned audio corpus. Finally, I discuss morphology as it relates to each of the four major Woods Cree parts of speech; nouns, pronouns, particles, and verbs. The data in this subsection will largely be drawn from Starks (1992) and Howse (1844), as well as from observations in the text corpus. As a complement to these data, I will also include various statements from an L1 consultant (Solomon Ratt of Stanley Mission). Concluding this subsection, I will provide templatic schematisations for the affix orderings of nouns and verbs. I do not discuss syntax in any great detail in this grammatical description; for a thorough overview of this topic, see Starks (1992).

2.1 Writing System

Woods Cree, in common with all other dialects of Cree, is a predominantly oral language, with English predominating as the written language across modern settlements. Most fluent speakers seldom, if ever, use the language in its written form. Despite this, like other dialects, written Woods Cree exists in a state of polygraphia. Ignoring early, aurally impressionistic transcriptions by European explorers, the earliest writing system for Woods Cree was that of Cree syllabics, an abugida-like syllabary initially developed in the 1830s by Methodist missionary James Evans for transcribing Ojibwe. With some modification, this writing system spread rapidly among Cree-speaking communities in the mid 19th century (Murdoch 1981), outpacing even the Christian missionaries responsible for its introduction.

	/i/	/i/	/o/	/o/	/a/	/a/	∅
/p/	ᐱ	ᐯ	ᐳ	ᐴ	ᐸ	ᐹ	ᑊ
/t/	ᑎ	ᑌ	ᑐ	ᑑ	ᑕ	ᑖ	ᐟ
/k/	ᑭ	ᑫ	ᑯ	ᑰ	ᑲ	ᑳ	ᐠ
/s/	ᓯ	ᓭ	ᓱ	ᓲ	ᓴ	ᓵ	ᐢ
/t͡ʃ/	ᒋ	ᒉ	ᒍ	ᒎ	ᒐ	ᒑ	ᐨ
/m/	ᒥ	ᒣ	ᒧ	ᒨ	ᒪ	ᒫ	ᒼ
/n/	ᓂ	ᓀ	ᓄ	ᓅ	ᓇ	ᓈ	ᐣ

Additional Characters	
/h/	ᐦ
/hk/	ᕽ
/l/	ᓬ
/r/	ᕒ

/j/	ᔨ	ᔦ	ᔪ	ᔫ	ᔭ	ᔮ	ᐩ
/ð/	ᖨ	ᖧ	ᖪ	ᖫ	ᖬ	ᖭ	ᖮ
/w/	ᐃᐧ	ᐁᐧ	ᐅᐧ	ᐆᐧ	ᐊᐧ	ᐋᐧ	ᐤ
∅	ᐃ	ᐁ	ᐅ	ᐆ	ᐊ	ᐋ	·

Table 6. Woods Cree syllabics (adapted from https://www.creedictionary.com/syllabics/woodland.php)

Syllabic texts dominate the Woods Cree written record in the late 19th and early 20th century, and continued to be widely used in personal correspondences well into the 20th century[18]. However, by the mid 20th century, various orthographies based on the Latin alphabet had supplanted the use of syllabics in academic and literary circles, following a general trend towards romanisation seen in many Western dialects of Cree (Bennett & Berry 1991:96); though syllabics are still used, most modern Cree texts are written in the Latin alphabet. By the 1970s, two distinct roman orthographies of Woods Cree had developed; these were the Standard Roman Orthography (or SRO) and the Colins Modified Orthography (or CMO), both of which remain in use to the present day.

2.1.1 Standard Roman Orthography

The Standard Roman Orthography (SRO) is an IPA-based romanisation initially developed throughout the 1970s for Plains Cree by Algonquianist linguists (Ellis 1973; Wolfart 1973; Pentland 1977) and Cree elders. The SRO was designed with the express intention of being extendable for cross-dialectal usage. Correspondingly, the Woods Cree adaptation of the SRO is identical to its Plains Cree counterpart (Okimāsis & Wolvengrey 2008), aside from its lack of the character ⟨ê⟩ (owing to Woods Cree lacking the relevant phoneme /eː/) and its inclusion of the digraph ⟨th⟩ (or, in some varieties, ⟨ð⟩) to represent the phoneme /ð/. In the corpus, 18 texts (constituting 59% of total tokens) used the character ⟨ð⟩, while 215 texts (constituting 41% of tokens) used ⟨th⟩. In addition to this, the SRO also sees variation in vowel length diacritics; most authors use circumflexes (e.g. ⟨â, î, ô⟩), while others use macrons (e.g. ⟨ā, ī, ō⟩). In the corpus, 212 texts (constituting 95.6% of total tokens) used circumflexes, compared to 21 texts (constituting 4.4% of total tokens) using macrons. Some authors also use acute accents for vowel length (e.g. ⟨á, í, ó⟩), but this does not occur in the text corpus. A full inventory of the characters used in the Woods Cree SRO, alongside their IPA equivalents, is given in Section 2.2.1.

The Standard Roman Orthography also has a complex set of rules regarding hyphenation, used to indicate certain morpheme boundaries and compounds. Person-number affixes (either on verbs or on nouns as possessors) are attached directly to the wordform without a hyphen, as are all

[18] For instance, in a text in the corpus (CLN-060422), Solomon Ratt describes using syllabics to write letters to his mother some time in the 1950s or 60s, including his (unsuccessful) attempts to transliterate English text into syllabics

suffixes. However, all prefixes which are not related to person-number agreement must be attached to the stem with a hyphen. If a person-number prefix attaches to a non-person-number prefix, it is attached to the non-person-number prefix without a hyphen, but the non-person-number prefix must then be attached to the stem (or to another prefix) with a hyphen. Consider the graphemic representation of the wordform ⟨*nikî-ati-nisitohtîn*⟩ ('I began understanding'); in this wordform, the 1SG prefix ⟨*ni*⟩ attaches directly to the past tense prefix ⟨*kî*⟩, which then attaches with a hyphen to the aspect prefix ⟨*ati*⟩, which also attaches with a hyphen to the stem ⟨*nisitoht*⟩. The 1SG suffix ⟨*în*⟩ then attaches to the stem ⟨*nisitoht*⟩ without a hyphen:

4) *nikî-ati-nisitohtîn*
ni-kî-ati-nisitoht-în
1SG.IND-PST-begin-understand(VTI)-1SG.IND
'I began understanding' (SM)

Although codified at length in Okimāsis & Wolvengrey (2008:16-21), hyphenation is, in practice, the least consistent aspect of the SRO between authors; even authors writing in otherwise identical SRO will often differ in the exact placement (or, indeed, the placement at all) of hyphens.

2.1.2 Colins Modified Orthography

While the SRO was merely adapted for use in Woods Cree, the other major romanisation, the Colins Modified Orthography (CMO), was developed specifically for it by two L1 speakers; namely, the eponymous Colin Charles and Keith Goulet, both of La Ronge. The pair developed the CMO in the 1970s with the express intention of offering it as a Woods Cree-specific alternative to the (at the time) newly-introduced SRO, on the basis of observations that elders and the "most fluent and knowledgeable speakers" in their community had "a difficult time" understanding the SRO (Charles et al. 2021). The CMO, therefore, was based on existing romanisations of Woods Cree, particularly those from early translations of religious texts. In Goulet's own words, for fluent Woods Cree speakers, the CMO was intended to be a system which was "already a part of their experience", being "more relatable" than the academically-developed SRO. Further, the CMO aimed to entirely avoid the use of diacritics in indicating vowel length, as Charles and Goulet found that such diacritics only introduced a "new layer of difficulty" in writing Woods Cree with a typewriter or computer. As a result, vowel length pairs are represented either with double characters or entirely different characters in the CMO: /a/ is represented as ⟨u⟩ while /aː/ is written as ⟨a⟩, short /i/ is written as ⟨i⟩ while /iː/ is written as ⟨e⟩, and short /o/ is written as ⟨o⟩ while /oː/ is written as ⟨oo⟩. Additionally, unlike the SRO, which represents /t͡s/~/t͡ʃ/ as ⟨c⟩, CMO represents this sound with the digraph ⟨ch⟩. An overview of correspondences between the SRO and CMO is given in Section 3.3.1. Finally, although hyphenation exists in the CMO, its usage is not codified, being instead governed largely

by authorial taste; in practice, it patterns similarly with its use in the SRO. A comparison of a sentence fragment containing all of these contrasts is provided below:

> SRO: *namwâc wîhtamawîw misi-kinîpikwa athisk kita-kôkîthiwa*
> CMO: numwach wehtumuwew misi-kinepikwu uthisk kitu-kookethiwa
> ‘He doesn’t tell Big Snake because she will dive’ (SM)

2.1.3 Contemporary Usage

Owing perhaps to its favour among academic circles, or its wealth of resources and widespread adoption in other Cree dialects, the SRO has ultimately proven the dominant orthography for Woods Cree, largely displacing the CMO even in La Ronge itself (Charles et al. 2021). However, the CMO continues to be used in some contemporary texts, perhaps most notably the recent digital re-release of the CCD (discussed in Section 1.5). In this investigation, I will exclusively be using the SRO (with the ⟨th⟩ digraph and circumflex length diacritics) unless specified otherwise. Given that the phonetic values of almost all characters in the SRO correspond with their IPA equivalents, I will provide additional IPA transcriptions only to highlight relevant phonetic distinctions which are not represented in the standard spelling.

2.2 Phonology

2.2.1 Phoneme Inventory

The phoneme inventory of Woods Cree, alongside the accompanying characters in the SRO, is as follows. A set of IPA charts showing common allophones may also be found in Appendix E:

	Bilabial	Dental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	/p/ ⟨p⟩		/t/ ⟨t⟩			/k/ ⟨k⟩	
Affricate			/t͡s~t͡ʃ/ ⟨c⟩				
Fricative		/ð/ ⟨th⟩	/s/ ⟨s⟩				/h/ ⟨h⟩
Nasal	/m/ ⟨m⟩		/n/ ⟨n⟩				
Approximant	/w/ ⟨w⟩				/j/ ⟨y⟩		

Table 7. Consonants of Woods Cree (/IPA/ and ⟨SRO⟩)

	Front	Mid	Back
High	/i/ ⟨i⟩ /iː/ ⟨î⟩		
Mid			/o/ ⟨o⟩ /oː/ ⟨ô⟩
Low		/a/ ⟨a⟩ /aː/ ⟨â⟩	

Table 8. Vowels of Woods Cree (/IPA/ and ⟨SRO⟩)

The affricate represented by ⟨c⟩ is largely in free variation between the alveolar [t͡s] and the alveopalatal [t͡ʃ], similar to Plains Cree (Okimāsis & Wolvengrey 2008:6) and Swampy Cree (Ellis 2000:xxviii). The stops /p/, /t/, and /k/ are universally non-aspirated, and /k/ is frequently realised as [g] intervocalically. [b] and [d] as voiced allophones of /p/ and /t/ are also attested intervocalically at South Indian Lake (Starks 1992:15), Stanley Mission, and La Ronge; however, Greensmith (1985:47,56) claims [b] and [d] to be absent in Pukatawagan. Even in communities where they do occur, [b] and [d] are much less commonly produced than [g]. Starks also reports [d͡ʒ] as an intervocalic allophone of [t͡s]~[t͡ʃ] at South Indian Lake:

5) a. *îkâ*
[iːˈgaː]
‘negative’
(SL) (Starks 1992:15)

b. *sîpâ*
[siːˈbaː]
‘under’

c. *ita*
[ɪˈda]
‘there’

d. *nicîmic*
[ˈnd͡ʒiːˌmɪt͡ʃ]
‘my younger sibling’

The semivowels /w/ and /j/ are often not fully distinct, and alternate with some regularity, particularly intervocalically. This alternation occurs in both Western and Eastern varieties, and appears to be at least somewhat lexically conditioned, although in South Indian Lake, Starks (1992:26) reports that /w/ systematically becomes /j/ when preceding a high front vowel, at least among younger speakers. A similar alternation exists between /h/ and /j/.

6) a. *nôhtâyîpan*
[noːhˈtaːjiːˌpʌn]
‘my late father’ (PK)

b. *nôhtâwîpan*
[noːhˈtaːwiːˌpʌn]
‘my late father’ (PK)

Note: Both uttered by the same speaker in the course of a single story

7) a. *mistahi*
[ˈmɪstʌˌhi]
‘a lot, really’ (PN)

b. *mistayi*
[ˈmɪstʌˌji]
‘a lot, really’ (PN)

Note: Both uttered by the same speaker in the course of a single story

Intervocalic semivowels are also prone to deletion, in which case the vowel preceding the semivowel is lengthened and the vowel following it coalesces into the first:

8) *iyako* > *îko*
[ˈijəˌko] [iːko]
‘that one’ (SL) (Starks 1992:28)

9) *ayamiw* > *âmiw*
[ˈajəˌmow] [ˈaːmow]
‘s/he speaks’ (SL) (Starks 1992:28)

There is also some degree of variation between the nasals /n/ and /m/, particularly at word boundaries. However, this alternation may also be lexically conditioned:

10) *nisiwanâtan* ~ *misiwanâtan*
[ˌnɪsɪˈwʌnaːˌtʌn] [ˌmɪsɪˈwʌnaːˌtʌn]
‘it is spoiled, it is destroyed’ (SL) ‘it is spoiled, it is destroyed’ (SM)

Greensmith (1985:101) and Starks (1992:27) also report instances of regressive nasal assimilation, with /n/ becoming [m] before /p/ and [ŋ] before /k/ in Pukatawagan and South Indian Lake. However, this process is not obligatory (see example 12):

11) a. *nika-nikamon*
[ŋkaˈnigaˌmon]
‘I will sing’ (PK)

b. *î-nipât*
[ˈiːmˌpaːt]
‘he sleeps’ (PK)

But

12) *ta-kî-nipâhtay*
[taˈgiːnpaːhˌtej]
‘he could have slept’ (PK)

At first glance, the Woods Cree vowel system appears to be a symmetrical one, with three short vowels (/a/, /i/, /o/) and three corresponding long vowels (/aː/, /iː/, /oː/). However, in practice, long and short vowels differ in quality as well as length; short /o/, for example, may be realised as [ʊ], short /i/ is often realised as [ɪ], and short /a/ is often realised as [ʌ]. The long vowel /iː/ and the sequences /ij/, /ih/, and /aji/ are sometimes realised as [eː], although it is “often very difficult” to distinguish between these segments in rapid speech (Greensmith 1985:97). The long vowel /oː/ and the sequences /ow/ and /iw/ are also sometimes realised as the high front vowel [u], possibly as an influence from English. Unstressed short vowels, particularly /i/ and /a/ are often reduced to /ə/, or deleted entirely. Correspondingly, speakers sometimes use unstressed /i/ and /a/ interchangeably (see example 15):

13) *mahihkan*
[ˈmaheh ˌkʌn]
‘wolf’ (PK)

14) *kisî-manitow*
[ˌkɪsiːˈmʌntu]
‘God’ (LR)

15) *kinîpik* ~ *kanîpik*
[ˈkɪniːˌpɪk] [ˈkʌniːˌpɪk]
‘a snake’ (PK) ‘a snake’ (PN)

Unlike /a/ and /i/, /o/ typically does not differ in quality in its short and long variants, aside from its allophone [ʊ]. Perhaps for this reason, it is often more difficult to ascertain the length of a given /o/ compared with other vowels, particularly stem-initially. For instance, when *o*-initial stems are given prefixes, the length of the initial /o/ appears to be in free variation:

16) *okimâw* > *nitôkimâm*
[ˈokɪˌmaːw] [nɪˈtoːkɪˌmaːm]
okimâw ni-t-okimâw-im
boss(NA) 1SG.POSS-LC-boss(NA)-POSS.THM
‘chief, boss’ (LR, SM, PK, SL) ‘my supervisor (PN)

But

17) *otokimâma*
[ˌotoˈkɪmaːˌmʌ]
o-t-okimâw-im-a
3SG.POSS-LC-boss(NA)-POSS.THM-OBV
‘his boss’ (PK)

Certain vowel-semivowel sequences can also obfuscate vowel length. /ij/ and /ow/ are the most prominent examples of this:

18) *maskihkiy*
[ˈmʌskihˌki(ː)j]
‘medicine’ (LR)

19) *wâpamow*
[ˈwaːpəˌmo(ː)w]
‘s/he sees him/herself’ (LR)

Overall, the phoneme inventory of Woods Cree is nearly identical to those of the other Western Cree dialects. barring three major exceptions. Firstly, and most conspicuously, Woods Cree possesses the phoneme /ð/, which is lacking in Plains Cree and Swampy Cree. As mentioned in Section 1.3, this /ð/ occurs as a reflex of the Proto-Algonquian *r, realised as /j/ in Plains Cree and as /n/ in Swampy Cree. This being said, there are also instances of the PrA segments *ç or *x being realised in modern Woods Cree as /ð/ (or its voiceless allophone [θ]). This is more common in Western Woods Cree than in the East, where it is typically realised as /h/; even in Western communities, *ç and *x being realised as /ð/ is more common among older speakers, with younger speakers increasingly realising these cases as /s/.

20) Proto-Algonquian *erenyiwa (‘person’)
Plains Cree: *iyiniw*
Woods Cree: *ithiniw*
Swampy Cree: *ininiw*
Moose Cree: *ililiw*[19]

21) Proto-Algonquian *wexpani (‘his/her lung’)
Plains Cree: *ohpan*
Woods Cree:
La Ronge:
othpan [oθpʌn] among older informants, *ospan* among others (Pentland 1978:109)
South Indian Lake:
ohpan
Pukatawagan:
ohpam
Swampy Cree: *ohpan*
Moose Cree: *ohpan*

Secondly, Woods Cree lacks the /e:/~/i:/ distinction present in southern Plains Cree[20], Swampy Cree, and Moose Cree. This levelling appears to be a fairly recent innovation; Howse reports a distinction between the two (see examples 23 and 24), and Greensmith (1985:93-94) reports some older speakers frequently producing [e:] in lexemes where [i:] would typically be expected.

22) Proto-Algonquian *eθkwe·wa (‘woman’)
Plains Cree: *iskwêw*
Woods Cree: *iskwîw*
Swampy Cree: *iskwêw*
Moose Cree: *iskwêw*

[19] Historically, the Moose Cree form of this word was *iliniw*; the second /l/ is the result of hypercorrection (Ellis 2004:324). A similar hypercorrection is reported in historical Woods Cree; in 1696, Henry Kelsey reports *ithithiw* as the Woods Cree form of this lexeme, rather than *ithiniw* (Pentland 1978:108)
[20] Northern Plains Cree has also merged /e:/ and /i:/; note, for example, the Southern Plains Cree minimal pair *wêpinam* (s/he throws it) and *wîpinam* (‘s/he soils it’), compared with the Northern Plains Cree *wîpinam* (s/he throws it) and *wîpinam* (‘s/he soils it’)

23) Néesh-oo
nîšo
[niːˈʃo]
‘two’ (UL) (Howse 1844:308)
Note: From PrA *nyi·šwi, compare with *nîso* [niːˈso] in modern Woods Cree and Plains Cree and *nîšo* [niːˈʃo] in Swampy and Moose Cree; all four (including historical Woods Cree) retain the PrA vowel /iː/

24) Náyw-oo
nêwo
[neːˈwo]
‘four’ (UL) (Howse 1844:308)
Note: From PrA *nye·wi, compare with modern Woods Cree *nîwo~nîyo* [niːˈwo]~[niːˈjo] and modern Plains, Swampy, and Moose Cree *nêwo* [neːˈwo]; modern Woods Cree shifts the PrA /eː/ to /iː/, while other Western Cree dialects and historical Woods Cree retain /eː/

[eː] is likewise occasionally reported in the speech of bilingual (or bidialectal) speakers of Woods Cree, likely as a borrowing, rather than an archaism:

25) *mistahi mâna kî-papâmôtêhowak ithiniwak*
[ˈmɪstaˌhi maːˈnʌ ˌkiːpʌˌpaːmoːˈteːhoˌwʌk ɪˈðɪnoˌwʌk]
‘the people used to travel around a lot’ (SM)
Note: Speaker also fluent in Plains Cree

Finally, Woods Cree has not preserved the Proto-Algonquian /s/~/ʃ/ distinction present in eastern dialects of Swampy Cree and Moose Cree (see example 26) (Oxford 2023a). Greensmith reports observing [ʃ] as an allophone of /s/ in Pukatawagan, occurring immediately preceding /tʃ/; however, I have found no evidence of a systematically allophonic /ʃ/ in Western Woods Cree.

26) a. *nakiskaw*
[ˈnʌkɪsˌkaw]
‘a short while’ (Swampy Cree)

b. *nakiškaw*
[ˈnʌkɪʃˌkaw]
‘Meet him/her!’ (Swampy Cree)

27) *pisci*
[pɪʃˈt͡ʃɪ]
‘by accident’ (PK) (Greensmith 1985:39)

This again appears to be a relatively recent development; in the late 17th century, the English fur trader Henry Kelsey reported the form *keeshque* for the modern Woods Cree *kîskwîw* (‘s/he is out of his/her mind’, compare to PrA *ki·waškwe) (Castel & Westfall 2001:xxxi), and Howse

reliably reports a small number of lexemes with ⟨sh⟩ where /s/ would be expected in contemporary speech. [ʃ] is also reliably attested in the corpus in the speech of a woman born in Southend, Saskatchewan in 1924, although here it seems to be an allophone of /ð/, rather than an archaism:

28) Péyshoohayoo
pêšiwêw
[ˈpeːʃɪˌweːw]
's/he brings him/her' (Howse 1844:41)
Note: Compare to *pêšiwêw* [peːʃɪweːw] in modern Swampy Cree and Moose Cree and *pîsiwîw* [piːsɪwiːw] in modern Woods Cree

29) *îkwânî tâpwî pîšis namwâc îkwa natonawîw*
[ˈigwaːˌniː taːˈpwiː piːˈʃis nʌmˈwaːt͡ʃ ˈiːkwa naˈtonoˌwiːw]
'so finally, he truly ceases to search for him' (Brightman 2007:178)
Note: Compare *pîšis* [piːʃis] to the expected form *pîthisk* [piːðisk]

There are also a small number of lexemes in which the reflex of the Proto-Algonquian *š is in free variation between /s/ and /t͡ʃ~t͡s/; however, this may simply be a phonological influence from modern Swampy Cree, rather than preserved palatalisation from PrA.

30) Proto-Algonquian *ešinlehkawete·wi ('it has such a name')

isîthikâtîw	~	*icîthikâtîw*
[ˌɪsiːˈðɪkaːˌtiːw]		[ˌɪt͡ʃiːˈðɪkaːˌtiːw]
'it is called (thus)' (PK)		'it is called (thus)' (PK)

Note: Compare to Swampy Cree *išinihkâtêw* [ˌɪʃɪˈniːhkaːˌteːw]

2.2.1.1 Marginal Phonemes

The phoneme /l/ is marginal in Woods Cree (Starks 1987:336), occurring chiefly in English and French loanwords. It is especially common in phonological adaptations of European names. Starks also notes /l/ as appearing in place of /ð/ in child-directed speech (see example 31c), however, having found no examples of CDS outside of South Indian Lake, I cannot speak to how widespread this phenomenon may be. Aside from this, however, the presence of /l/ is attested across the Woods Cree dialect area:

31) a. *ciliwî*
[ˈt͡ʃɪlɪˌwiː]
'killdeer (Ln. *Charadrius vociferus*)'[21] (LR)

b. *omakalakisa*
[oˌmʌkəˈlʌksˌə]
'his/her mukluks' (SL)

[21] Likely a borrowing of the corresponding English term, although *ocâpiskâkanîsiw* (lit. 'the little creature with a scarf') is also attested at La Ronge

c. *kâla*
[kaːˈlʌ]
‘Don’t’ (SL)
Note: Only occurs in child-directed speech; typically, this is *kâtha* [kaːˈðʌ]

In addition to /l/, /r/ (which varies allophonically between [ɹ] and [ɾ]) is also attested in Woods Cree (Starks and Ballard 1994:104-5), however, its distribution is even more marginal than /l/. Thus far, I have found it to be restricted entirely to European personal names. /f/ has a similarly restrictive distribution, also occurring in a loanword for ‘coffee’, however, critically, all of the loanwords in which it occurs also have attested forms with [p]. As such, it is difficult to disentangle whether these segments are indeed marginal phonemes in Woods Cree or simply byproducts of Woods Cree speakers, most of whom also speak English, pronouncing English proper names in a more authentically English fashion. My inclination is towards the latter, as I have yet to identify any minimal pairs involving these marginal segments.

32) a. *sîril*
[siːˈrɪl]
‘Cyrille’ (PK)
Note: *sithîl* is also attested

b. *lariwa*
[ˈlariˌwə]
‘Larry’ (SL) (Starks & Ballard 2005)
Note: *hâthî* (‘Harry’) is also attested

33) a. *filâp*
[fɪˈlaːp]
‘Philip’ (PK)
Note: *pilâp* is also attested

b. *kâwfî*
[kaːwˈfiː]
‘coffee’ (PK)
Note: *kôpiy* is also attested

2.2.2 Syllable Structure

The canonical structure of Woods Cree syllables is as follows:

(C(w)) V ((s, h)C)

According to this schematisation (adapted from Greensmith (1985:30)), a Woods Cree syllable consists of an optional consonant onset, followed by an optional /w/, then an obligatory single vowel nucleus, followed by an optional /s/ or /h/ and an optional coda consonant.

2.2.2.1 Onset

All consonants may occur in the onset position, except for /h/ and /j/, which only occur as onsets in a small number of discourse particles and onomatopoeia:

34) a. *hay-hay*
[hʌj hʌj]
‘thank you’ (SM)

b. *hâw, ôta api*
[haːw oːˈta ʌˈpɪ]
‘Okay, sit here’ (SM)

c. *hay ya, hay ya, hay ya, paskawâpisimowin ninikamohtân*
[hʌˈjə hʌˈjə hʌˈjə ˌpʌskaˌwaːpɪˈsɪmoˌwɪn ˌnɪnɪˈgamohˌtaːn]
‘Hey-ya, Hey-ya, Hey-ya, the Shut-Eye dance I sing’ (SM)

The only consonant clusters allowed in onsets are /pw/, /tw/, /kw/, and /mw/:

35) a. *pwâwîw*
[pwaːˈwiːw]
‘s/he is pregnant’

b. *kwîskitî*
[ˈkwiːskɪˌtiː]
‘the other side’

c. *twîhow*
[twiːˈhow]
‘s/he lands’

d. *mwîstas*
[mwiːsˈtəs]
‘after’

2.2.2.2 Nucleus

Woods Cree has no diphthongs; a single vowel always constitutes the syllabic nucleus. However, this vowel may be highly reduced or elided, leading to a variety of non-phonemic clusters. When in an unstressed position, short /i/ and short /a/ are particularly susceptible to elision and mutual substitution:

36) a. *ômânisis*
[oˈmaː**ns**ˌɪs]
‘moose-calf’ (PK)
(Greensmith 1985:123-4)

î-nîpinithik
[ˌiːniːˈpɪ**n**ˌ**ð**ɪk]
‘when it was summer’ (PK)

paskwatinâhk
[paskˈwa**tn**aː]
‘on the bare land’ (PK)

2.2.2.3 Coda

Any singleton consonant is permitted in coda position, along with the clusters /hp/, /ht/, /hk, /hc/, /sk/, and /hð/, although /hð/ is uncommon (Starks & Ballard 2003:3):

37) a. *sihkihp*
[sihˈkihp]
‘helldiver (Ln. *Podiceps auritus*)’ (LR)

b. *âtiht*
[aːˈtiht]
‘some’ (LR)

c. *âtihk*
[aːˈtihk]
‘caribou (Ln. *Rangifer tarandus*)’ (LR)

d. *anohc*
[ʌˈnoht͡ʃ]
‘today’ (LR)

e. *amisk*
[ʌˈmɪsk]
‘beaver (Ln. *Castor canadensis*)’ (LR)

f. *wîhth*
[wiːhð]
‘Name him/her!’ (LR)

Consonant clusters of the form *Cw* can also occur underlyingly in the word-final position, however, in these instances, they are reduced to *C*. The existence of this underlying /w/ is only revealed in the presence of suffixes:

38) a. *otâpânâsk*
[oˈta:pa:ˌna:sk]
otâpânâskw
car(NA)
'car' (SM)

b. *otâpânâskwak*
[ˌota:ˈpa:na:sˌkwʌk]
otâpânâskw-ak
car(NA)-PL
'cars' (PK)

c. *otâpânâskohk*
[ˌota:ˈpa:na:sˌkohk]
otâpânâskw-ihk
car(NA)-LOC
'in/at/to the car' (PK)[22]

d. *ocâpânâskos*
[ˌot͡ʃa:ˈpa:na:sˌkos]
ocâpânâskw-is
car(NA)-DIM
'a small car' (PK)[23]

Longer clusters consisting of three consonants do occasionally occur (i.e. /skw/ in example 38b above); however, these are only seen word medially, occurring as the result of adjacent codas and onsets:

39) *ta**htw**âw*
[tahtˈwa:w]
'every time' (LR)
Note:([taht])$_{\sigma}$([wa:w])$_{\sigma}$

2.2.3 Elision and Reduction

Haplology is extremely common, occurring when an unstressed *CV* syllable immediately precedes a (stressed) syllable with the same onset consonant; in these instances, the unstressed syllable is deleted, and the onset of the remaining syllable is (optionally) slightly lengthened:

40) *ninanâskomon*
[nɪˈn(ˑ)a:skoˌmon]
'I am thankful' (SM)

A related process to this cluster reduction is the deletion of word initial /ki/, particularly when onset of the following syllable is /k/, /s/, or /t/. Similarly, as mentioned in Section 1.4, the Conjunct marker *kâ-* is frequently reduced to simply *â-* in South Indian Lake and Pukatawagan;

[22] Attested form in the corpus is *pôsotâpânâskohk* (from *pôsotâpânâsk*(*w*) ('bus'))
[23] Attested form in the corpus is *ocâpânâskosa*

so frequently, in fact, that Greensmith (1985:110) notes one of her consultants in Pukatawagan "does not hear the difference" between the two:

41) a. *kika-kî-itohtân*
[ˌkɪkəˈgi:tohˌta:n]
'you can go' (PK)

b. *ka-kî-itohtân*
[kʌˈgi:tohˌta:n]
'you can go' (PK)

42) a. *kâ-wâpahtamân*
[ˌka:wa:ˈpʌhtəˌma:n]
'I see it' (SM, LR, PN)

b. *â-wâpahtamân*
[ˌa:wa:ˈpʌhtəˌma:n]
'I see it' (SL, PK)

In general, unstressed short vowels are highly susceptible to deletion, particularly /i/. This process of deletion is so ubiquitous that it is (occasionally) given a dedicated orthographic representation, with the character ⟨'⟩ used in place of the deleted vowel[24]:

43) *tân'si*
[ˈta:nsɪ]
'hello' (SM)
Note: The orthographic form *tânsi*, with no trace of the unstressed /i/, is also common

The stops /p/, /t/, and particularly /k/ can also be deleted intervocalically in rapid speech , as can /ð/, which is also occasionally realised as [t]~[d] intervocalically (Greensmith 1985:109-13). Intervocalic /k/ is also reported to be occasionally realised as [ɣ] or [ʔ] at Pukatawagan (see examples 46 and 47), with this intervocalic environment even extending across word boundaries. However, these intervocalic allophones of /k/ are still less common than [g], which is near ubiquitous.

44) *kisîthiniw*
[kɪˈsi:t(ɪ)now]
'old man' (SM)

45) *poko*
[po:]
'only' (SL) (Starks 1992:26)

46) *â-takopathiyâhk*
[a:ˌtaɣoˈpaðiˌja:]
'we arrived' (PK) (Greensmith 1985:106)

47) *îtokî*
[ˈi:toˌʔi:]
'I guess' (PK) (Greensmith 1985:106)

Greensmith also reports [β] as an allophone of the cluster /pw/, although she only attests this in the speech of one consultant.

[24] Although commonly used, this convention is not universally accepted; in Okimāsis & Wolvengrey(2008;21-4), for instance, an entire section is dedicated to discouraging the use of apostrophes to mark unstressed vowels.

48) *â-sîpwîpitahk*
[ˌa:si:ˈβi:pɪˌta]
‘he started it’ (PK) (Greensmith 1985:104)

Consonant clusters are sometimes reduced to singleton consonants when occurring in word-final position. For instance, /sk/ is sometimes reduced to /s/ (see example 29), and /h/-initial clusters are frequently reduced to /h/ word-finally (see Section 2.2.5).

Vowel-semivowel sequences are frequently reduced to single vowels, with adjacent /w/ segments often rounding /i/ or /a/ to /o/. For instance, the sequences *aw*, *awi*, *wi*, and *iw* may all be reduced to *o* (or *ow*). The sequences *iy*, *hi*, and *ih* are also frequently reduced to *î*.

49) a. *nîhithaw*
[ˈni:hɪˌðaw]
‘Cree person’ (LR, PK)

b. *nîhithow*
[ˈni:hɪˌðow]
‘Cree person’ (SM)

c. *nîthow*
[ˈni:ðow]
‘Cree person’ (PN)

50) a. *kimiwan*
[ˈkɪmɪˌwʌn]
‘it rains’ (KN)
(Torkornoo et al. 2021)

b. *kimowan*
[ˈkɪmoˌwʌn]
‘it rains’ (LR)

When the sequence *iwi* is reduced, it is sometimes rounded to /o/, and sometimes kept unrounded as /i/ or /i:/. Relatedly, the sequence *îwî* may be reduced to /i:/. These reductions may even occur across word boundaries:

51) *îkotî wîpac*
[ˌi:goˈti:pʌt͡ʃ]
‘there suddenly (PK) (Greensmith 1985:73)

The sequence *wîw* may be reduced to [oyu], and *wa* may also occasionally be reduced to [u]. /o/ may also become /u/ when in the sequences /jo/ (see example 52) and /ow/ (see example 14):

52) *niyo*
[niˈju]
‘four’ (PK) (Greensmith 1985:84)

Relatedly, in semivowel final stems, the addition of any suffixes beginning in short /i/ causes the stem-final semivowel to be deleted, the /i/ in the suffix to deleted, and the resultant stem-final vowel to be compensatorily lengthened:

53) a. *mîskanâs*
[ˈmi:skʌˌna:s]
mîskanaw-is
road(NI)-DIM
‘little road’ (SM)

b. *waskwâhk*
[wʌsˈkwa:hk]
waskway-ihk
birch(NA)-LOC
‘on the birch’ (SM)

Eastern varieties in particular also show a tendency to reduce *Cw* clusters to singleton consonants in rapid speech. Eastern varieties also occasionally delete word final /w/, particularly in the syllable *wâw*, reducing it to *wâ* (see, for example Section 2.3.3.1.1). Both phenomena are demonstrated in the following example:

54) *mihcîtwâw*
[ˈmihtʃ͡i:ˌta:]
‘many times’ (PK) (Greensmith 1985:90)

2.2.4 Epenthesis

In instances where affixation would result in a sequence of multiple different vowels, linking consonants are inserted to avoid clustering. For example, when the 1SG, 2SG, or 3SG prefixes *ni-*, *ki-*, and *o-* are added to a vowel-initial stem, the linking consonant /t/ is obligatorily inserted (although this is also sometimes realised as /h/ or /j/)

55) a. *nitapin*
[ˈnɪtəˌpɪn]
‘I sit’ (PK) (Greensmith 1985b)

b. *niyapin*
[ˈnɪjəˌpɪn]
‘I sit’ (PK) (Greensmith 1985b)

In certain possessed nouns, the linking consonant /t/ is sometimes realised as [t͡s~t͡ʃ]; however, whether this change is phonologically motivated, in a state of free variation, or simply a lexicalised idiosyncrasy of certain nouns remains unclear:

56) a. *ocahcâpiya*
[ˌot͡ʃahˈca:piˌjə]
‘his bow’ (PK)

b. *nicawâsimis*
[ˌnɪt͡ʃəˈwa:sɪˌmɪs]
‘my child’ (SM)

Aside from these person-number prefixes, most (vowel-final) affixes trigger an optional /h/ or /j/ as a linking consonant; however, the pronunciation of this consonant is often very slight and difficult to distinguish:

57) *kîhtî-ayak*
[ki:hˈti:hɪˌjʌk]
‘elders’ (Prince Albert, SK)

These linking consonants are typically reserved to avoid clusters of multiple vowels of differing quality. By contrast, when affixation would cause a sequence of two vowels of identical quality, the vowels in question may alternatively collapse into a single long vowel:

58) *tânitî kâ-atoskîyin?*
[ˈtaːnti: kaːˈtoskiːˌjən]
‘Where do you work’ (KN) (Torkornoo et al. 2021)

2.2.5 Pre-Aspirated Consonants

Consonant clusters beginning with /h/, sometimes referred to as ‘pre-aspirated consonants’ (e.g. Starks & Ballard 2003; Schmirler 2016) have the effect of obscuring vowel length contrasts (Pentland 1978:112-3), with the length of any given vowel being difficult to distinguish when followed by a *hC* cluster. This phenomenon is not unique to Woods Cree, being noted in both Plains Cree (Wolvengrey 2011a:126) and Swampy Cree (Pentland 1978:113).

Word-final cluster reduction is a phenomenon present across Woods Cree, but more prevalent in its Eastern varieties. Greensmith (1985:126-7) reports that the word-final elision of the cluster /hk/ is a categorical rule in Pukatawagan, with (optional) word medial /hk/ deletion also being attested. Similarly Starks (2003) reports that word final reduction of /hk and /hp/ to /h/ is ubiquitous among younger speakers in South Indian Lake, although not categorical among older speakers, with non-reduced forms still regularly being produced by them. Speakers in Brochet are noted by Brightman (2007:167) as reducing word final /hk/ to [x], as are speakers from other communities near the Eastern-Western isogloss such as Sandy Bay, Southend, and Sturgeon Landing. This /hk/ > [x] shift is also attested in Swampy Cree (Pentland 1978:113). Speakers from Sandy Bay, Southend, and Sturgeon Landing are also noted as often reducing pre-aspirated consonants to singleton consonants and compensatorily lengthening the preceding vowel (e.g. realising *pîwinahki* (‘s/he scatters it’) as *pîwinâki*). By contrast, in Western communities such as Stanley Mission, La Ronge, and Pelican Narrows, pre-aspirated consonants are more commonly retained, with /hk/ typically being preserved word-finally (although simplification to /h/ does still occasionally occur). This generally mirrors the behaviour of word-final /hk/ in Plains Cree (Schmirler 2016). When this cluster reduction causes /h/ to occur word-finally, its pronunciation is very slight and often difficult to perceive; Starks (1992:16) describes it as “whispered”. /h/ seldom occurs word-finally elsewhere.

59)

a.	*î-takosih*	b.	*nîpisîhkopâh*
	[iːˈtʌkoˌsih]		[niːpˈsiːhkoˌpaːh]
	î-takosi-hk		nîpisîhkopâw-ihk
	CNJ-arrive(VAI)-3SG.CNJ		willow.grove(NI)-LOC
	‘s/he arrives’ (PN)		‘in the willow grove’ (SM)

2.2.6 Stress

Like other Cree dialects, Woods Cree stress is fixed and predictable, rather than lexically determined. The system in Woods Cree is similar to that which has been reported in other Western Cree dialects (Wolvengrey 2022): in two syllable words, stress falls on the final syllable, and in words of three syllables or more, primary stress falls on the antepenult, with secondary stress occurring on alternating syllables to the left and right.

60)	a.	*piyak* [piˈjʌk] ‘one’ (LR)	b.	*môswa* [moːsˈwʌ] ‘moose’ (LR)
61)	a.	*takwâkin* [ˈtʌkwaːˌkɪn] ‘it is autumn’ (SM)	b.	*atâmihk* [ˈʌtaːˌmihk] ‘underneath’ (SM)
62)	a.	*mînisihtakâw* [ˌmiːnɪˈsihtəˌkaːw] ‘there is an orchard’ (SM)	b.	*masinahikanâhtik* [ˌmʌsɪˌnahɪˈkʌnaːhˌtɪk] ‘pencil’ (PK)

Monosyllables do exist, but are exceedingly uncommon, being limited to a small number of loanwords, imperatives, and discourse particles:

63)	a.	*tî* [tiː] ‘tea’ (PK)	b.	*sôp* [soːp] ‘soap’ (PK)	c.	*mîth* [miːð] ‘Give it to him/her!’[25] (SM)

Stress and vowel length are entirely uncorrelated; short vowels may be stressed and long vowels may be unstressed:

64)	a.	*sâwanohk* [ˈsaːwəˌnohk] ‘south’ (SM)	b.	*mitâtaht* [ˈmɪtaːˌtaht] ‘ten’ (SM)

Even when deleted, unstressed short vowels still impact how stress assignment proceeds on the lexical unit as a whole. In example 65, for instance, even though the unstressed /i/ is deleted,

[25] In South Indian Lake, an /a/ may be optionally inserted at the end of imperative forms such as this to avoid monosyllables (e.g. *mîtha)*

primary stress is still assigned to the first syllable *tân*, not the final syllable *si*, as would be expected for a typical disyllabic word.[26]:

65) *tânisi*
[ˈtaːnsɪ]
'hello' (SM)

The exact nature of stress in Western Cree dialects is an elusive topic; Schmirler & Arnhold (2017) report that stress in Plains Cree primarily manifests through rises in pitch and intensity, but also report that these pitch contours are governed principally by a phrasal pitch accent system, with this phrase-level system overriding the previously reported lexical system. Although, impressionistically, Woods Cree stress also appears to primarily impact pitch and intensity, I have investigated this matter at only a very cursory level, and leave the nature of the acoustics of Woods Cree stress as an open question to future researchers.

2.3 Morphology

2.3.1 Nouns

Woods Cree nouns fill largely the same semantic and syntactic niche as nouns in English, representing concrete physical objects and locations, as well as abstract concepts and states. Nominal morphology is governed by a binary system of animacy-based grammatical gender, whereby all nouns must be either animate (NA) or inanimate (NI). This animacy value impacts various aspects of the noun's potential morphosyntactic distribution:

66) a. *ithiniw* (NA)
'person' (LR)

b. *masinahikan* (NI)
'book'(LR)

The grammatical animacy of nouns in Woods Cree generally overlaps with the real-world animacy of the referent in question; however, this is not always the case (see Section 2.3.1.2):

67) a. *asiniy* (NA)
'stone' (LR)

b. *ospwâkan* (NA)
'pipe' (LR)

Animacy is typically unmarked on singular nouns; instead, the primary indicators of a noun's animacy are its available choice of demonstratives (see also Section 2.3.2.2), the verb stems for which it may act as an argument (see Section 2.3.4.7), and its plural marker (NA *-ak*, NI *-a*):

[26] It is for precisely this reason that the SRO represents unstressed short vowels, even in words where they are almost never pronounced (Okimāsis & Wolvengrey 2008:21-4)

68) a. *ôma masinahikan misâw*

ôma	masinahikan-∅	misâ-w
DEM.PROX.INAN.SG	book(NI)-SG	be.big(VII)-3SG.IND

‘this book is big’ (NC)

b. *ôho masinahikana misâwa*

ôho	masinahikan-a	misâ-wa
DEM.PROX.INAN.PL	book(NI)-PL	be.big(VII)-3PL.IND

‘these books are big’ (NC)

c. *awa atim miskitiw*

awa	atimw-∅	miskiti-w
DEM.PROX.ANIM.SG	dog(NA)-SG	be.big(VAI)-3SG.IND

‘this dog is big’ (NC)

d. *ôko atimwak miskitiwak*

ôko	atimw-ak	miskiti-wak
DEM.PROX.ANIM.PL	dog(NA)-PL	be.big(VAI)-3PL.IND

‘these dogs are big’ (NC)

This animacy marking system is reduced in comparison with that of Proto-Algonquian, which explicitly marked animacy and singular number on nouns through affixation. In modern Woods Cree, these animacy-number suffixes are retained on only a small number nouns in order to avoid monosyllables:

69) a. *maskwa*
maskw-a
[bear-ANIM.SG](NA)
‘bear’ (LR)

b. *wâwi*
wâw-i
[egg-INAN.SG](NI)
‘egg' (LR)

2.3.1.1 Inflectional Category Suffixes

In addition to differences in demonstratives, verb stems, and plural marking, there are a number of inflectional suffixes whose assignment is restricted by the animacy of the noun to which they attach. These are the obviative, the locative, the vocative, the distributive, and the absentative. These mutually exclusive[27] suffixes, particularly locative and vocative, have been referred to by some authors as case suffixes (Starks 1992:30; Wolvengrey 2011a:39), given that they serve to explicitly mark the grammatical function of their head noun. However, I would caution against the using the term ‘case’ to describe these suffixes; none of these suffixes can be applied to nouns universally (except for the Obviative, which is only universal among animate nouns), their

[27] The absentative is the exception to this mutual exclusivity, and can occur with the obviative; see Section 2.3.1.1.5

distributions are often syntactically unpredictable, and many of these suffixes, particularly the vocative, absentative, and distributive, are not fully productive. As such, I will refer to these suffixes with the more generic term ‘inflectional category suffixes’ going forward.

2.3.1.1.1 Obviative

Woods Cree, as with all other Algonquian languages, has two distinct third-persons; the proximate (3) and the obviative (3`). The proximate is used when referring to the (third-person) entity which is of the greatest relevance, prominence, or importance to the discourse at hand; all other, less prominent third-person entities in the discourse are encoded with the obviative.

70) *“âpakosîs ôma nîtha,” itwîw awa âpakosîs. ikwa ati-ây-itwahîw otôtîma*

âpakosîs-∅	ôma	nîtha	itwî-w	awa
mouse(NA)-PROX	FOC	1SG	say(VAI)-3SG.IND	DEM.PROX.ANIM.SG

âpakosîs-∅	ikwa	ati-ây-itwah-îw
mouse(NA)-PROX	and	start-RDPLW-point(VTA)-3SG>3`.IND

o-tôtîm-a
3SG.POSS-friend(NDA)-OBV
‘“Mouse is who I am,” says Mouse and he points to his friends’ (SM)

As the above example demonstrates, proximate nouns are morphologically unmarked, whereas obviative nouns are marked, receiving the suffix *-a*. Obviative nouns also lose their number distinction; consider the following example, where the obviative form *otôtîma* is used to refer to one, single friend, as opposed to multiple friends (as it does in <u>example 70</u>):

71) *pîkwîthihtam awa âpakosîs athisk piyak otôtîma î-ati-nakatât*

pîkwîthiht-am	awa	âpakosîs-∅	athisk	piyak
be.sad(VTI)-3SG.IND	DEM.PROX.ANIM.SG	mouse(NA)-PROX	because	one

o-tôtîm-a	î-ati-nakat-ât
3SG.POSS-friend(NDA)-OBV	CNJ-start-abandon-3SG>3`.CNJ

‘Mouse is sad because he has left one of his friends.’ (SM)

Broadly speaking, obviation is a category only overtly marked on grammatically animate nouns, with grammatically inanimate nouns always being morphologically unmarked for this category (although they may still act as obviatives syntactically (see, for instance, <u>Sections 2.3.2.2</u> and <u>2.3.4.7.1</u>)). However, in the corpus, there are some examples of Eastern Woods Cree speakers overtly marking inanimate nouns as obviative using the suffix, *-ithiw*, the same suffix used for most inanimate obviative demonstratives (see <u>Section 2.3.2.2</u>). Whether these cases represent isolated innovations or a more coherent, productive set remains an open question.

72) a. *kî-âpacihtâw pasastihikan akwa tîhtipiwinithiw*

kî-âpacihtâ-w	pasastihikan	akwa	tîhtipiwin-ithiw
PST-use(VTI)-3SG.IND	whip(NI)	and	chair(NI)-OBV

‘He used a whip and chair’ (PK)
Note: Discussing an animal tamer training big cats

b. *mîth arnî masinahikanithiw*

mîth	arnî	masinahikan-ithiw
give(VTA)2SG>3SG.IMP.IMM	Ernie	book(NI)-OBV

‘Give Ernie a book!’ (PK)

All transitive verbs with multiple animate third-person arguments are required to mark at least one of their arguments as obviative. Owing to their generally being of lesser discursive relevance, the obviative-marked argument is typically the object:

73) a. *iskwîw wâpamîw nâpîwa*

iskwîw	wâpam-îw	nâpîw-a
woman(NA)	see(VTA)-3SG>3`.IND	man(NA)-OBV

‘the woman sees a man/men’ (NC)

b. *iskwîwa wâpamîw nâpîw*

iskwîw-a	wâpam-îw	nâpîw
woman(NA)-OBV	see(VTA)-3SG>3`.IND	man(NA)

‘the man sees a woman/women’ (NC)

Similarly, when possessed by a third-person, animate nouns must be marked as obviative:

74) a. *nitôtîm*
ni-tôtîm
1SG.POSS-friend(NDA)
‘my friend’ (NC)

b. *otôtîma*
o-tôtîm-a
3SG.POSS-friend(NDA)-OBV
‘his/her friend(s)’ (NC)

c. **otôtîm*
o-tôtîm
3SG.POSS-friend(NDA)
‘*his/her friend’ (NC)

2.3.1.1.2 Locative

The locative suffix *-ihk* (broadly glossed as ‘at, in, on, etc.’ (Wolfart 1973:31)) is used to indicate location; Starks (1992:45) describes this as ‘chang[ing] a noun to an adverb of place’:

75) *ikota matâwisihk pîhow awa kisîtiniw*

ikota	matâwis-ihk	pîho-w	awa	kisîtiniw
there	clearing(NI)-LOC	wait(VAI)-3SG.IND	DEM.PROX.ANIM.SG	old.man(NA)

‘There in the clearing the old man waited.’ (SM)

76) *otinam cîstatâyâpiy omaskimotihk*
otin-am cîstatâyâpiy o-maskimot-ihk
take(VTI)-3SG.IND sinew(NI) 3SG.POSS-bag(NI)-LOC
'he takes sinew from his bag' (PK)

Although it is entirely grammatical without them, the locative is often paired with adpositions or other locational modifiers so as to clarify the relevant spatial aspects of the referent in question:

77) *misawâc ispimihk mistikohk nitakosân*
misawâc ispimihk mistikw-ihk ni-tihtakosî-n
in.any.case up tree(NA) 1SG.IND-climb(VAI)-1SG.IND
'In any case I am up a tree' (SM)

Unlike the obviative, which is constrained by grammatical animacy, the locative suffix is instead constrained by biological animacy; that is, the locative may be applied to any noun which refers to a biologically non-living entity (regardless of its grammatical animacy), and cannot be applied to any biologically living entity. Consider the following examples:

78) *nikâwîpan âhkosîwikamikohk î-kî-ayât*
ni-kâwiy-ipan âhkosîwikamikw-ihk î-kî-ayâ-t
1SG.POSS-mother(NDA)-ABST hospital(NI)-LOC CNJ-PST-be.there(VAI)-3SG.CNJ
'my late mother was in the hospital' (SM)

79) *kâwi kâ-pôsiyân nitotâpânâskohk*
kâwi kâ-pôsi-yân ni-otâpânâskw-ihk
back CNJ-board(VAI)-1SG.CNJ 1SG.POSS-car(NA)-LOC
'I got back in my car' (SM)

80) **niwîpinîn ithinîhk*
ni-wîpin-în ithiniw-ihk
1SG.IND-throw(VTI)-1SG.IND person(NA)-LOC
*'I throw it at the person' (NC)

The locative is equally grammatical with *âhkosîwikamik* ('hospital'), a grammatically and biologically inanimate noun, as it is with *otâpânâsk* ('car'), a grammatically animate, but biologically inanimate noun. By contrast, it is ungrammatical with *ithiniw* ('person'), a grammatically animate and biologically animate noun. This contrast between biological animacy and grammatical animacy is discussed further in Section 2.3.1.2.

The locative suffix is extremely common in toponyms, often becoming fully lexicalised with them (see example 81). Many adpositional particles also feature this suffix (see example 82), although here too, it has become lexicalised and inseparable from the particle itself:

81) *mistahi-sâkahikanihk*
mistahi-sâkahikan-ihk
big-lake(NI)-LOC
'La Ronge, SK (lit. 'at the big lake')' (LR) (Torkornoo et al. 2021)

82) a. *pîhcâyihk*
'inside, in the middle' (LR, SM, PK)

b. *ispimihk*
'up, up above, high up' (LR,SM,PK)

2.3.1.1.3 Distributive

The distributive suffix *-inâhk*, which may be glossed as 'in the land of …' (Wolfart 1973:31) or 'among the …' (Van Eijk & Collette 2020:852), is used to indicate a location replete with, or closely associated with, the noun in question. Like the locative, it is over-represented in toponyms:

83) a. *kimosôm-pwâtinâhk*
ki-mosôm-pwât-inâhk
2SG.POSS-grandfather(NDA)-Sioux-DSTR
'Deschambault Lake' (LR, PN)[28]

b. *kimosôminâ*(*hk*)
ki-mosôm-inâhk
2SG.POSS-grandfather(NDA)-DSTR
'Old Man's Bay' (PK)

c. *ithinînâhk*
ithiniw-inâhk
person(NA)-DSTR
'a native settlement' (PK)

The distributive is much less common than the locative; only eight total examples could be found in the corpus, most of which were highly lexicalised. Seven out of these eight instances[29] were found affixed to nouns which are both grammatically and biologically animate. This distribution largely resembles the pattern seen for the equivalent morpheme in Plains Cree (Van Eijk & Collette 2020:857), where it is slightly more common.

[28] https://learncree.ca/northern-saskatchewan-place-names
[29] The exception being *nitasikaninâninâhk* ('in our stockings'),which is grammatically, but not biologically, animate.

2.3.1.1.4 Vocative

The vocative, used to identify a noun as being directly and personally addressed, manifests in two ways. Forming the singular vocative generally involves eliding the final consonant(s) in a noun and lengthening the resultant word-final vowel (although some irregular singular vocatives also exist (see example 87b)). The plural vocative, by contrast, is always formed with the suffix *-itik*:

84)	a.	*nikosis*	b.	*nikosî*
		ni-kosis		ni-kosî
		1SG.POSS-son(NDA)		1SG.POSS-son(NDA).VOC.SG
		‘my son’ (SM, LR, PN, PK, SL)		‘My son!’ (SL)
85)	a.	*nicânis*	b.	*nicânisitik*
		ni-cânis		ni-cânis-itik
		1SG.POSS-daughter(NDA)		1SG.POSS-daughter(NDA)-VOC.PL
		‘my daughter’ (PK, SL)		‘My daughters!’ (SL)

Connotationally, the vocative is considered a polite form of address, implying respect to the addressee; indeed, it is has even been referred to by some authors as the ‘honorific’ (Ratt 2014:xii; Van Eijk & Collette 2020)[30]. Perhaps relatedly, the vocative is attested in the corpus almost exclusively with various kinship terms:

86) *kahkithaw niwâhkômâkanitik, nipakosîthimon kita-mithopathihikoyîk*
kahkithaw ni-wâhkômâkan-itik, ni-pakosîthimo-n
all 1SG.POSS-relative-VOC.PL 1SG.IND-hope(VAI)-1SG.IND
kita-mithopathih-iko-yîk
FUT-cause.to.fare.well(VTA)-INAN-2PL.CNJ
“All my relations, I hope you have good fortune” (SM)

The only non-kinship terms attested in the corpus with the vocative are terms related to God (in the Christian sense); a speaker may begin a prayer, for example, with the exhortation *kisîmanitô* (‘Oh God!’, from *kisîmanitow* (‘God’), see example 87a). Additionally, the form *nôhtâ* (‘My father!’, from *nôhtâwiy* (‘my father’)) is also sometimes used as an interjection of surprise or shock. This may be a calque of the English ‘(oh) my God’, as *nôhtâwiy* is commonly used as a term for the Christian Supreme Being.

[30]A personal anecdote may do well to illustrate this point; when I asked an L1 Plains Cree speaker if the vocative form *nitêmitik* (‘Hark, my dogs!’) was grammatical, she replied that, although she had never heard it, she could imagine that an individual with close personal affection to their dogs might say this. When asked if she would ever say this to her own dogs, she responded “No, I don’t even give them names!”

87) a. *kisîmanitô, mithin wanaskîwin*

kisîmanitô	mith-in	wanaskîwin
God(NA)	give(VTA)-2SG>1SG.IMP.IMM	serenity(NI)

'God, grant me serenity' (SM)

b. *tâpwî nôhtâ aspin kîtahtawîth kâ-sakîcipathit*

tâpwî	n-ôhtâ	aspin	kîtahtawîth	kâ-sakîcipathit
truly	1SG.POSS-father(NDA).VOC	then	presently	CNJ-convulse(VAI)-3SG.CNJ

'Then, my God, all of the sudden she went into an [epileptic] seizure' (SL)

Contemporary usage of vocative forms (particularly singular vocative forms) is in decline across the Woods Cree dialect area. While the modern vocatives are limited to kinship terms, historically, vocative forms existed for a much wider variety of nouns:

88) a. Eskwayw-etíck
iskwîtik
iskwîw-itik
woman(NA)-VOC.PL
'Women!' (Howse 1844:183)

b. I`etíck
ayitik
ay-itik
living.being(NA)-VOC.PL
'Such a one!' (Howse 1844:183)

Even in their limited, modern distribution, vocative forms are increasingly being substituted in favour of unmodified singular or plural forms. Starks (1992:45), for example, records the following instance of a single speaker addressing his grandfather directly as both *nimosô* (the vocative form) and *nimosôm* (the non-vocative form) in the course of a single interaction:

89) a. *nimosô, î-takosinân*

ni-mosô	î-takosin-ân
1SG.POSS-grandfather(NDA).VOC	CNJ-arrive(VAI)-1SG.CNJ

"Grandfather, I'm home." (SL)

b. *nimosôm, kîkwân ôma kâ-ôh-mâtowin*

ni-mosôm	kîkwân	ôma	kâ-ôh-mâto-win
1SG.POSS-grandfather(NDA)	what	FOC	CNJ-for-cry(VAI)-2SG.CNJ

"Grandfather, why are you crying?" (SL)

This shift away from vocative usage is not unique to Woods Cree (similar patterns have been noted in Swampy Cree (Ellis 2000:xxxvi) and Plains Cree (Wolvengrey 2011a:39); however, based on available corpus data, the shift does appear more advanced in Woods Cree than its neighbours.

2.3.1.1.5 Absentative

The absentative (or preterital) suffix, *-ipan*, is used in Woods Cree to indicate that the referent of a noun is deceased. Correspondingly, the absentative can only be applied to biologically animate nouns, specifically those referring to humans. Although, in terms of corpus frequency, this suffix has a strong collocational preference for kinship terms, it can be attached to any noun referring to a human being, including to personal names:

90) a. *nohtâwîpan*
n-ohtâwiy-ipan
1SG.POSS-father(NDA)-ABST
'my late father' (SM)

b. *nistîsipan*
ni-stîs-ipan
1SG.POSS-older.brother(NDA)-ABST
'my late older brother' (PK)

c. *ana môniyâskwîpan*
ana môniyâskwîw-ipan
DEM.MED.ANIM.SG White.woman(NA)-ABST
'this late White woman' (PK)

d. *cânipan*
cân-ipan
John-ABST
'the late John' (PK)

I have found no examples in the corpus of the absentative co-occurring with nouns referring to non-human animals; as such, it appears to have a similar distribution to the vocative and distributive, at least pertaining to the animacy of its referents.

The absentative is attested throughout the geographic distribution of Woods Cree, as well as in other Western Cree dialects. Its Plains Cree equivalent is virtually identical, applying to kinship terms, proper names, and some general terms for humans (e.g. *nôcokwêsipan* ('late old woman', from *nôcokwêsis* ('old woman')) (Wolfart & Ahenakew 2000). The Swampy Cree absentative serves a similar function, but seems to have a wider distribution; for example, it can attach (albeit 'facetiously' (Ellis 2004:386)) to ruined inanimate nouns:

91) *nicîmânipan*
ni-cîmân-ipan
1SG.POSS-canoe(NI)-ABST
'my old canoe' (Swampy Cree) (Ellis 2004:386)

Of the inflectional category suffixes discussed thus far, the absentative is by far the least inflectional in character; indeed, it would not be unreasonable to analyse the modern Woods Cree

absentative as simply a derivational affix with a somewhat unusual morphosyntactic distribution. However, I have included it here due to its inflectional origins; the modern day nominal absentative is a remnant of a larger, historical set of past tense paradigms known as the Preterit (see Section 2.3.4.8.3). Historically, affixes related to the Preterit could be used to mark past tense meanings on verbs, nouns, and even pronouns.

92) Sákeh uké pun
sâkihakipan
sâkih-akipan
love(VTA)-1SG>3SG.CNJ.PRET
'If I loved him' (Howse 1844:301)

All three of these functions are preserved in modern Moose Cree:

93) a. *nimosômipan*
ni-mosôm-ipan
1SG.POSS-grandfather(NDA)-ABST
'my late grandfather' (Moose Cree) (James 1991:282)

b. *apipan têtâwitak môčitak*
api-pan têtâwitak môčitak
sit(VAI)-3SG.IND.PRET in.the.middle on.the.floor
'he would sit in the middle of the floor' (Moose Cree) (James 1991:284)

c. *awênipan otohôsima čân*
awêna-ipan ot-ohôs-im-a čân
who-PRET 3SG.POSS-horse(NA)-POSS.THM-OBV John
'John's horse was gone!' (Moose Cree) (James 1991:282)

Based on this morphological heritage, the Woods Cree absentative (at least, historically) might be interpreted more abstractly as a nominal tense marker. According to this interpretation, example 90a (*nohtâwîpan*), rather than meaning 'my deceased father' might be interpreted as 'my past father' or 'my former father', with the affixal semantics shifting over time from 'past/former' to 'deceased' by virtue of context. This progression is supported by historical evidence of the absentative being used with inanimate referents in Plains Cree, where it indicated that the referent either no longer existed, was presently absent, was lost, or was no longer used, a diverse variety of potential nominal interpretations of past tense:

94) Ni mokkumâniban
nimôhkomânipan
ni-môhkomân-ipan
1SG.POSS-knife(NI)-ABST
‘mon ancien couteau’ (Plains Cree) (Lacombe 1874:section 2, page 18)
Note: ‘my former knife, my old knife (which I no longer use/possess)’

Precisely how the Woods Cree absentative came to be reduced in distribution to only sentient animate nouns is not clear, and requires further historical research. However, as it exists in the modern language, the absentative, much like the distributive and vocative, occupies a curious, semi-productive morphosyntactic niche somewhere between inflection and derivation.

2.3.1.2 Grammatical Animacy versus Biological Animacy

As demonstrated throughout the previous sections, grammatical and biological animacy do not universally overlap in Woods Cree. Instead, they appear to operate as two, semi-independent systems, at least as it relates to determining which inflectional category suffixes a given noun may receive. As such, one may think of Woods Cree animacy as operating as a graded spectrum, rather than a dichotomy. Within this spectrum, nouns may be entirely inanimate (biologically and grammatically), ‘counter-intuitively’ animate (that is, biologically inanimate, but grammatically animate), or entirely animate (biologically animate and grammatically animate). This final category may then be further subdivided into sentient and non-sentient:

		Example	Vocative	Absentative	Distributive	Obviative	Locative
NA	Sentient	*ithiniw* ‘person’	✓	✓	✓	✓	
	Grammatically and Biologically Animate	*sîsîp* ‘duck’			?	✓	
	‘Counter Intuitively’ Animate	*asiniy* ‘stone’			?	✓	✓
NI	Grammatically and Biologically Inanimate	*akohp* ‘blanket’				?	✓

Table 9. A comparison of the distribution of inflectional category suffixes among nouns based on their biological and grammatical animacy. ✓ indicates the morpheme in question may co-occur with nouns of the specified type, while ? indicates that this co-occurrence is rare, marginal, or otherwise non-systematic.

Within this animacy spectrum, the subdivision of 'counter-intuitively' animate nouns deserves further explication. 'Counter-intuitively' animate nouns are a regular occurrence in other dialects of Cree (see, for instance, Van Eijk & Collette 2020), and, although highly lexicalised, tend to have at least one of the several qualities described in the list below. Qualities 1 and 2 on this list are strong, predictive determinants of counter-intuitive animacy; most or all biologically inanimate nouns which possess these qualities are grammatically animate. The remainder are general diagnostic trends; some, but not all, nouns with these qualities are grammatically animate.

1) Capacity for self-propelled motion or growth

otâpânâsk	*naskasiy*
'car' (SM)	'my fingernail' (SM)

2) Involvement or association with spirits; spiritual "power" (Van Eijk & Collette 2020:841)

ospwâkan	*ahcahk*
'pipe' (LR)	'soul, spirit' (LR)

3) Association with winter and winter survival

asâm	*kôna*
'snowshoe' (LR)	'snow' (LR)

But not

miskotâkay	*astotin*
'jacket' (LR)	'hat' (LR)

4) Body parts or garments which naturally occur in pairs

mitihtihkosiw	*astis*
'kidney'[31] (LR)	'glove' (LR)

But not

mihpan	*maskisin*
'lung' (LR)	'shoe' (LR)

Although these animacy guidelines can explain most counter-intuitively animate nouns, a minority remain inexplicable:

95) a. *îmihkwân*	b. *sôniyâw*	c. *athôskan*
'spoon' (LR)	'money' (LR)	'raspberry'[32] (LR)

[31] *mitihtihkosiw* sometimes also occurs as an inanimate noun.
[32] Other berries, such as *ithinimin* ('blueberry'), *kaskitîmin* ('blackberry'), *môsomin* ('bush cranberry' (Lt. *Viburnum trilobum*)) are typically inanimate.

A further factor which can promote a biologically inanimate noun to grammatical animacy is analogy; if a biologically inanimate noun has a referent which is closely analogous to that of an established grammatically animate noun, the former noun can 'absorb' the animacy of the latter, even if it lacks the qualities which made the latter noun animate to begin with. Consider the following example:

96) *nitatâwîstamâson piyakwâyihk sikarîtisak*
ni-t-atâwîstamâso-n piyakwâyihk sikarîtis-ak
1SG.IND-LC-buy.for.oneself-1SG.IND one.unit cigarette(NA)-PL
'I bought myself a carton of cigarettes' (SM)

The loanword *sikarîtis* ('cigarette') is grammatically animate, as evidenced by its plural marker, despite lacking any of the qualities associated with counter-intuitive animacy; cigarettes are incapable of growth or locomotion, have no inherent spiritual associations, are irrelevant to winter survival, and are not a body part. Rather, *sikarîtis* appears to be borrowing its animacy from two existing animate nouns whose real-world referents are similar to that of *sikarîtis*, *ospwâkan* ('pipe') and *cistîmâw* ('tobacco'), despite lacking the qualities that make *ospwâkan* and *cistîmâw* grammatically animate (namely, their association with spirits and ritual). A very similar form of animacy by analogy is attested in Plains Cree (Dacanay & Schmirler 2022).

Although grammatical animacy is a lexically determined quality of all Woods Cree nouns, it would be inaccurate to describe the animacy value of any given noun as entirely static. Grammatically inanimate nouns may be dynamically promoted to animacy without any overt formal alteration when discursively felicitous, and demoted once this discursive felicity lapses. Take the following example from *cihcipiscikwân* (The Rolling Head), a well-known legend in which a man decapitates his adulterous wife only for her severed head to regain consciousness and murderously pursue their children in revenge:

97) *ispî î-ati-pahkihtihk mohcihk isi tîpwâtam omiyaw anima mistikwân.*
ispî **î-ati-pahkihti-hk** mohcihk isi tîpwât-am
when **CNJ-start-fall(VII)-3SG** on.the.ground thus shout(VTI)-3SG
o-miyaw **anima** mi-stikwân
3SG.POSS-body(NI) **DEM.MED.INAN.SG** X.POSS-head(NDI)
'As the Head tumbled to the ground it yelled at the Trunk.' (SM)

mistikwân ('head') is typically a grammatically inanimate noun; this much is proven here by its use of inanimate demonstratives and verb stems. However, later in the story, once the head

begins to roll of its own volition, it transitions mid-sentence to being an animate noun with no formal changes,

98) *ikwa mihkiwâhpihk anima mistikwân, cihcipiscikwân awa, ati-tâh-tihtipîpathiw î-kâh-kakwîcimât âpacihcikanisa tânitî î-kî-itohtîthit otawâsimisa.*

ikwa	mihkiwâhp-ihk	**anima**	mi-stikwân	cihcipiscikwân
and	lodge(NI)-LOC	**DEM.MED.INAN.SG**	X.POSS-head(NDI)	Rolling.Head

awa	**ati-tah-tihtipîpathi-w**	**î-kâh-kakwîcim-ât**	
DEM.PROX.ANIM.SG	**start-RDPLS-roll(VAI)-3SG.IND**	**CNJ-RDPLS-ask(VTA)-3SG>3`.CNJ**	
âpacihcikanisa	tânitî	î-kî-itohtî-thit	otawâsimisa.
tool(NA)-DIM-OBV	where	CNJ-PST-go(VAI)-3`.CNJ	3SG.POSS.child-POSS-DIM-OBV

'Meanwhile, inside the lodge the severed head, this is now Rolling Head, rolls about asking the utensils where her children had gone.' (SM)

This sentence also demonstrates a second animacy promotion, with *âpacihcikan* ('tool, utensil'), usually grammatically inanimate, being treated as animate due to their role in the story as sentient creatures.

In addition to spontaneous animacy promotions such as these, there are also a number of animacy-based minimal pairs; that is, pairs of formally identical lexemes distinguished in meaning solely by their animacy. For example, when used as an NA, *mistik* means 'tree', while when used as an NI, it means 'stick':

99) a. *miscikosa î-âpacihtât ta-nawacît ikota kotawânihk.*

mistikw-is-a	î-âpacihtâ-t	ta-nawacî-t
stick(NI)-DIM-PL	CNJ-use(VTI)-3SG.CNJ	FUT-roast(VAI)-3SG.CNJ

ikota	kotawân-ihk.
there	fireplace(NA)-LOC

'He used sticks to roast them over the fire' (SM)

b. [*mi*]*toni kî-mâh-misikiticik mistikwak*

mitoni	kî-mâh-misikiticik	mistikw-ak
very	PST-RDPLS-be.big(VAI)-3PL.IND	tree(NA)-PL

'The trees were huge' (PK)

There also exist a number of lexemes whose animacy varies between speakers and communities with no semantic change. For example, *sôniyâw* ('money') is typically an NA, but is also occasionally produced as an NI, with no apparent systematic contrast. Other animacy alternations

are geographically mediated; while *pimithâkan* (‘airplane’) is consistently produced as an NA at Stanley Mission, it is only ever attested as an NI at Pukatawagan:

100) a. *môtha kîkwan sôniyâw nikî-ohci-ayân*
môtha kîkwan sôniyâw ni-kî-ohci-ayâ-n
nothing money(NI) 1SG.IND-PST-from-have(VTI)-1SG.IND
‘We did not have any money’ (PK)

b. *sôniyâw isa kâ-otiniht*
sôniyâw isa kâ-otin-iht
money(NA) reportedly CNJ-take(VTA)-X>3SG.CNJ
‘The money was taken’ (PK)

c. *â-wâpahtamâ îkota pimithâkana*
(k)â-wâpahtam-â(hk) îkota pimithâkan-a
CNJ-see(VTI)-1PL.CNJ there airplane(NI)-PL
‘We saw planes there” (PK)

d. *ikwa ikota ohci pimithâkanisa nipôsinân*
ikwa ikota ohci pimithâkan-is-a ni-pôsi-nân
and there from airplane(NA)-DIM-OBV 1SG.IND-board(VAI)-1PL.IND
‘and from there we’d board a small bush plane’ (SM)

2.3.1.3 Possession

Noun possession in Woods Cree is expressed morphologically using a coordinated set of prefixes and suffixes, which attach directly to the possessed noun:

1SG	ni-		
2SG	ki-		
3SG	o-		(-a)
1PL	ni-		-inân
12PL	ki		-inaw
2PL	ki-		-iwâw
3PL	o-		-iwâw(a)
3′	o-		-thiw(a)

Table 10. Possessive paradigm for nouns

Possessive affixes do not differ on the basis of animacy; however, animate nouns with third-person possession must also be obviative (as indicated in parentheses in Table 10). Starks (1992:46) reports that nouns with plural possessors cannot be inflected for the locative at South Indian Lake; however, this combination is attested elsewhere in the corpus:

101) a. *â(y)-isinâkosit otôniwâhk wîtha*

(k)â-isinâkosi-t	o-tôn-iwâw-ihk	wîtha
CNJ-look.like(VAI)-3SG.CNJ	3PL.POSS-mouth(NDI)-3PL.POSS-LOC	FOC

'that's what it looked like in their mouths' (PK)

When attached to vowel-initial stems, possessive prefixes typically trigger a linking consonant /t/, /h/, or /j/ to avoid an illegal vowel sequence. However, for some stems (particularly those which begin with /o/), the first and second-person possessive prefixes may instead be optionally shortened to *n-* and *k-* respectively:

102) a. *nôkimâminân*

ni-okimâw-im-inân

1PL.POSS-chief(NA)-POSS.THM-1PL.POSS

'our (excl.) chief' (PK) (Greensmith 1985:77)

b. *kônîkânîmiwâwak*

ki-ônîkânîw-im-iwâw-ak

2PL.POSS-leader(NA)-POSS.THM-2PL.POSS-PL

'your (pl.) leaders' (PK) (Greensmith 1985:77)

Possession may be optionally emphasised through the use of personal pronouns in addition to the possessive affixes; however, possessive morphology remains obligatory in these instances:

103) *kâ, îyako nîtha nitîhim.*

kâ	îyako	nîtha	ni-tî-h-im
indeed	PRON.INAN.SG	1SG	1SG.POSS-tea-LC-POSS.THM

'That's right, this is my tea.' (PK)

2.3.1.3.1 Dependent Nouns

A substantial set of Woods Cree nouns are inalienable, and must always occur with possessive morphology; in general Algonquianist terminology, these are referred to as Dependent nouns. Dependent nouns may be animate (NDA) or inanimate (NDI), and are generally either kinship

terms or body-part terms[33], although a small set of highly personal material possessions are also afforded dependent status:

104)	a.	*nikâwiy*	b.	*kicihciy*	c.	*oskotâkay*
		ni-kâwiy		ki-cihciy		o-skotâkay
		1SG.POSS-mother(NDA)		2SG.POSS-hand(NDI)		3SG.POSS-jacket(NDI)
		'my mother' (LR, SM)		'your hand' (LR)		'his/her jacket' (LR)

In addition to affixes for first, second, and third-person possession, some dependent nouns can also be marked for unspecified (or indefinite) possession using the prefix *mi-*

105)	a.	*micihciy*	b.	*miskîsikohkâna*
		mi-cihciy		mi-skîsikohkân-a
		X.POSS-hand(NDI)		X.POSS-eyeglass(NDI)-PL
		'a hand, someone's hand' (LR)		'a pair of glasses, someone's glasses' (LR)

This unspecified possessor prefix has a much more restricted distribution than the other possessive affixes. Generally, *mi-* may be attached to dependent nouns for material possessions (e.g. *miskotâkay* 'a jacket') and human body parts (e.g. *micihciy* ('a hand')), but not body parts which are only found on non-human animals[34]. Its use with kinship terms is somewhat unclear; in the corpus, no examples of kinship terms with with the *mi-* possessor are attested, with unspecified possession instead being accomplished with a separate paradigm:

106)	a.	*nikâwiy*	b.	*okâwîmâw*
		ni-kâwiy		o-kâwiy-imâw
		1SG.POSS-mother(NDA)		X.POSS-mother(NDA)-X.POSS
		'my mother' (SM)		'a mother' (SM)

However, in the *Colin Charles Cree Dictionary*, several examples of kinship terms with *mi-* are provided as entries, with the Howse grammar also containing such examples:

107) a. Me-gáuwee
mikâwiy
mi-kâwiy
X.POSS-mother(NDA)
'a mother' (UL) (Howse 1844:245)
Note: Also attested contemporaneously in La Ronge

[33] One notable exception to inalienability in body part terms lies in expletive interjections, which are often clipped dependent nouns for lewd body-part terms (i.e. *takay!* - 'penis' (expected form *mitakay*), *cisk!* - 'posterior' (expected form *micisk*)

[34] Although *misoy* ('his/her tail') is listed in the CCD, and *mitahtahtakwanih* ('on the wings') is attested in Pukatawagan

b. *mistîs*
mi-stîs
X.POSS-mother(NDA)
'an older brother' (LR)

c. *mitânis*
mi-tânis
X.POSS-daughter(NDA)
'a daughter' (LR)

By contrast, in contemporary Plains Cree (Muehlbauer 2007:173) and Swampy Cree (Cenerini 2014:5), *mi-* typically does not combine with kinship terms, mirroring my observations from the Woods Cree corpus. This is also the opinion of my consultant, who did not recognise any of the proposed *mi-* kinship terms from the CCD (Solomon Ratt, personal communication via Arok Wolvengrey). In any case, the aversion to the *mi-* possessor with kinship terms seems to be a relatively recent phenomenon; in historical Western Cree dialects, *mi-* kinship terms are attested (i.e. Watkins 1865:323).

Assuming the distribution seen in the corpus (and described by my consultant), the unspecified possessor may only occur on biologically inanimate dependent nouns, regardless of their grammatical animacy. However, I leave the matter of determining the extent of the *mi-* unspecified possessor in contemporary Woods Cree to future field research.

The semantics of the unspecified possessor when used with body-part terms differ somewhat between Woods Cree and its neighbours. While in Plains Cree, the unspecified possessor is generally used for body parts when the possessor is generic, unknown, or even simply "irrelevant" (Wolfart 1973:15), the Woods Cree unspecified possessor is more often used for body parts which are "detached (Brightman 1985:355), either being severed (and thus no longer having an inalienable possessor) or belonging to a specific and relevant, but unseen or absent, referent (although both sets of meanings are possible in both dialects):

108) a. *ê-apihkâtahk mêscakâsa*
ê-apihkât-ahk m-êstakay-is-a
CNJ-braid(VTI)-3SG.CNJ X.POSS-hair(NDI)-DIM-PL
's/he is braiding hair' (Plains Cree)[35]

b. *wîpaha mistakaya*
wîpah-a m-istakay-a
sweep(VTI)-2SG.IMP.IMM X.POSS-hair(NDI)-PL
'Sweep up the hair'
Context: 'Directive to child after family haircuts' (BR) (Brightman 1985)

[35] https://speech-db.altlab.app/maskwacis/segment/14881

c. *wâpahtamwak micihciy*
wâpahtam-wak mi-cihciy
see(VTI)-3PL.IND X.POSS-hand(NDI)
'they saw someone's hand' (BR) (Brightman 1985)
Note: From a story in which a group of people hide in a hole while a cannibalistic monster gropes around the area with its hands in an attempt to find them

2.3.1.3.2 Possessive Theme *-im*

When possessed, some non-Dependent Woods Cree nouns take the suffix *-im*, described by Wolfart (1973:28) as the "possessive theme sign". This theme sign is not unique to Woods Cree, and occurs in other Western Cree dialects (Wolfart 1973:28; Ellis 2015:83), where its function and distribution have been described as "subject to a great deal of irregularity" (Wolfart 1973:28). Likewise, in Woods Cree, this suffix is erratic in its occurrence, and difficult to assign any consistent semantic or syntactic function. It occurs with both animate and inanimate nouns, although (in the corpus) it is markedly more frequent among the former than the latter. It is obligatory with some nouns, but for most, its presence is optional, and engenders no reliable semantic shift. It has been postulated to be a focal marker of some kind, or an indicator of increased or emphasised affiliation between the possessor and the referent (Howse 1844:184; Katherine Schmirler, personal correspondence), however, any attempts to outline a coherent usage case for the morpheme, other than that it must co-occur with possession, are invariably met with manifold counter-examples:

109) a. *otamiskoma*
o-t-amiskw-im-a
3SG.POSS-LC-beaver(NA)-POSS.THM-OBV
'his/her beavers' (PK)

b. *onôcokwîsîma*
o-nôcokwîsiw-im-a
3SG.POSS-old.lady(NA)-POSS.THM-OBV
'his/her old lady' (PK)

c. *kîspin wanikiskisiyâni nika-wanihâw nitahcahk*
kîspin wanikiskisi-yâni ni-ka-wanih-âw
if forget(VAI)-1SG.SUBJ 1SG>3SG.IND-FUT-lose(VTA)-1SG>3SG.IND
ni-t-ahcahkw
1SG.POSS-LC-soul(NA)
'If I forget, I will lose my soul' (SM)

d. *otahcahkoma nitahâw pikwîta kîkway kâ-osihtâthân*
o-t-ahcahkw-im-a ni-t-ah-âw
3SG.POSS-LC-soul(NA)-POSS.THM-OBV 1SG>3SG.IND-LC-put(VTA)-1SG>3SG.IND

pikwîta kîkway kâ-osihtâ-thân[36]
everywhere PRON.INAN.SG CNJ-create(VTI)-1SG.CNJ
'her soul is woven into everything I create' (PN)

The position of the *-im* theme relative to other suffixes is also a matter of irregularity. Typically, *-im* occurs immediately following the noun stem; however, there are occurrences in the corpus of it occurring after the diminutive suffix *-is*. As with other aspects of this morpheme, it is unclear if its position relative to noun stem has any impact on its function or meaning.

110) a. *ikwâni mîthîw ana iskwîw otoskawâsisima âyâs okâwîthiwa*
ikwâni mîth-îw ana iskwîw
then give(VTA)-3SG>3`.IND DEM.MED.ANIM.SG woman(NA)
o-t-oskawâs-is-im-a âyâs o-kâwî-thiw-a
3SG.POSS-baby(NA)-DIM-POSS.THM-OBV âyâs 3`.POSS-mother(NDA)-3`.POSS-OBV
'The woman gives the baby to Âyâs's mother.' (SM)

2.3.1.4 Diminutive

Diminutive nouns are extremely common in Woods Cree, and are formed through two simultaneous processes. These are the addition of the suffix *-is* and the substitution of all /t/ segments in the stem[37] for /t͡s~t͡ʃ/:

111) *mistikowat* > *miscikowacis*
'wooden box, wooden trunk' (PK) 'little wooden box' (PK)

Multiple diminutive suffixes can be added to the same noun stem. This is particularly common with semivowel-final stems; indeed, some diminutive semivowel-final stems are only attested in the corpus with multiple diminutive suffixes. Typically, these 'double diminutives' have their diminutive semantics augmented or emphasised, although this augmentation may be less noticeable in certain lexicalised cases.

112) a. *sîpîsis*
sîpiy-is-is
river(NI)-DIM-DIM
'creek' (PK, SM)
Note: Compare *sîpîs* ('little river') (Arok Wolvengrey, personal communication)

[36] I believe this may have been a speech error caused by hypercorrection, as the speaker who produced this sentence is also fluent in Plains Cree. The expected (Woods Cree) form here would be *kâ-osihtâyân*
[37] /t/ segments in inflectional affixes are usually not subject to this substitution, although linking consonants for possessive prefixes are sometimes affected (e.g. *nicawâsimis* ('my child'))

b. *iskwîsis*
iskwîw-is-is
woman(NA)-DIM-DIM
'girl' (PK, SM)
Note: The single diminutive *iskwîs* ('small (adult) woman') is not attested in corpus, although the cognate form *iskwês* is seen in Plains Cree.[38]

In double diminutive forms such as these, the vowel in the second *-is* suffix is sometimes lengthened:

113)	*sâkahikan*	>	*sâkahikanis*	>	*sâkahikanisîs*
	'lake' (LR)		'pond' (LR)		'little pond' (PK, SM)

This lengthening is not obligatory, and contrastive pairs with and without lengthening exist. For instance, for one speaker (Solomon Ratt, personal correspondence) *sâkahikanisis*, the double diminutive of *sâkahikan* without lengthening, is the general term for a small pond, whereas the double diminutive with lengthening, *sâkahikanisîs*, is the name of a specific body of water near Little Stanley Rapids, a freshwater strait some kilometres northeast of Stanley Mission.

As one might expect, the diminutive is typically used to indicate small size, limited quantity or extent, incomplete maturation, or affection (although, in its extremely high frequency, its use is occasionally somewhat capricious[39]). However, for reasons which I have been unable to ascertain, there exists a small subset of nouns which do not inherently possess any of these qualities which are permanently lexicalised in the diminutive:

114) a. *simâkanis*
'police officer' (PK, SM, LR)
Note: Seemingly diminutivised from a now-archaic term for 'spear', *simâkan* (Lacombe 1874:593)

b. *ayiwinis*
'clothes'(LR)
Note: compare to *ayiwin* ('article of clothing') in Plains Cree (Wolvengrey 2011b)

[38] Relatedly, *nâpîs* ('boy'), the (single) diminutive of *nâpîw* ('man'), is attested in the corpus, but is notably less common than the double diminutive *nâpîsis* (3 instances compared to 19 instances respectively)
[39]Consider for example, the following sentence from the corpus: *sîwîpitîwak opaspihowîwa West Coast Trails ohci wîtha kahkithaw iyako akihcâsowinis î-kî-mîthikawiyâhk*. ('they called the West Coast Trail rescue number which we all had'). Quite why *akihcâsowinis*, the diminutive of *akihtâsowin*, the word for a number or single digit, is used for a phone number containing multiple digits, I cannot say.

2.3.1.4.1 Diminutive Sound Symbolism

Before proceeding, it would be prudent to grant the diminutive /t/ > /t͡s/~/t͡ʃ/ sound shift a modicum further attention. This shift, referred to in the literature as Diminutive Consonant Symbolism (DCS) (Pentland 1974), is not categorically required, and diminutive forms are regularly produced without it (although these purely affixal diminutives are less common than diminutives with both DCS and the diminutive suffix):

115)	a.	*pôtâcikanis*	b.	*piskitis*
		pôtâcikan-is		piskit-is
		whistle(NI)-DIM		biscuit(NA)-DIM
		'small whistle' (SM)		'biscuit' (SM)

By contrast, diminutive nouns formed using only DCS (and not the diminutive suffix) have a much more restricted distribution, occurring only in kinship terms and affectionate nicknames. However, it can be difficult in some instances to distinguish this phenomenon from allophonic variation between /t/ and /t͡s/~/t͡ʃ/ (see Section 2.2.4); consider, for instance, example 116b, which could plausibly be either:

116) a. *âcihk*
'Little Caribou (pet name for a dog)' (UL) (Pentland 1974:241)
Note: Stem *âtihk*, expected diminutive *âcihkos*[40]

b. *nicânis*
'my daughter' (PK)
Note: Stem *nitânis*, expected diminutive *nicânisis*

In describing Plains Cree, Wolfart (1973:80) mentions that DCS may be applied to entire sentences, imparting them with an "overly sweet, feminine, [or] pitiful" tone; however, the Woods Cree corpus contains no such examples of sentence-level DCS[41]. Starks (1992:18) also notes DCS to be common in child-directed speech, although this may simply be due to diminutives in general (including those with the diminutive suffix) to be common in child-directed speech. This is attested, for instance, in Swampy Cree (Melnychuk 2003), as well as many other typologically unrelated languages (Alderete & Kochetov 2016). However, lacking any lengthy segments of child-directed speech in the corpus, this claim is impossible to investigate here without substantial fieldwork.

[40] *âcihkos* (or *acihkwas*) is also attested as (human) personal name at Pukatawagan
[41] Incidentally, I have similarly found no examples of this phenomenon in Plains Cree corpora, and my Plains Cree consultants have categorically rejected sentence-level DCS if not accompanied by the requisite diminutive suffixes. However, I freely admit that my field work in the topic has been fairly preliminary, and would keenly encourage further research on the matter.

2.3.1.5 Deverbal Nominalisation

Given the reputation of Cree dialects as having a "preference for saying thing[s] with verbs' (Ellis 2000:xxxii), it should come as little surprise that a significant portion of nouns in Woods Cree are simply nominalisations of verb stems. Some of the most common deverbal nominalisation patterns are:

-win

A general-purpose nominaliser which may combine with virtually any verb to create a noun with the meaning 'the act of *V*', although 'the place where *V* occurs', 'the result of *V*', and 'instrument with which *V* is performed' are also occasionally seen:

117) a. *pimâtisiwin*
pimâtisi-win
[live(VAI)-NOM](NI)
'life' (SM)

b. *kaskihtâwin*
kaskihtâ-win
[be.able(VTI)-NOM](NI)
'ability, competence' (UL)

c. *kapîsiwin*
kapîsi-win
[reside(VAI)-NOM](NI)
'campsite' (SM, PK, LR)

d. *mikîwin*
mikî-win
[give(VAI)-NOM](NI)
'a gift' (LR)

e. *tîhtapiwin*
tîhtapi-win
[sit(VAI)-NOM](NI)
'chair' (LR, PK)

-ikan

Nouns formed with *-ikan* generally have the meaning 'instrument associated with V' or 'manufactured item exemplifying quality V', however, 'individual who is characteristic of V' is also an attested pattern of meaning.

118) a. *cîkahikan*
cîkah-ikan
[chop(VTI)-NOM](NI)
'axe' (SM, PN, LR, PK)

b. *âpacihcikan*
âpaciht-ikan
[use(VTI)-NOM](NI)
'tool' (SM, LR, PK)

c. *akwanahikan*
akwanah-ikan
[cover(VTI)-NOM](NI)
'a tarp, a covering' (PK, SM)

d. *kihtimikan*
kihtim-ikan
[be.lazy(VAI)-NOM](NA)
'a lazy person' (LR)

-âkan

The precise semantics of *-âkan* are variable; typically, nouns formed with this suffix have a meaning broadly definable as 'goal of V' (Cook & Muehlbauer 2010:7), although some *-âkan* nouns also fit the instrumentative role usually taken by *-ikan*:

119)	a.	*wîcîwâkan*	b.	*môhkotâkan*
		wîcîw-âkan		môhkot-âkan
		[accompany(VTA)-NOM](NA)		[whittle(VTI)-NOM](NI)
		'friend' (SM, LR, BR)		'a planer, a whittling tool' (LR, PK)

o-

The *o-* prefix is attached to verbs in their 3SG Independent form to create animate nouns with the general semantic pattern of '(habitual) performer of V':

120)	a.	*omanâpitîpicikîw*	b.	*owanihikîw*
		o-manâpitîpicikî-w		o-wanihikî-w
		[NOM-pull.teeth(VAI)-3SG.IND](NA)		[NOM-trap(VAI)-3SG.IND](NA)
		'dentist' (LR)		'a trapper' (PK, LR)

Collectively, nouns formed using these five affixes alone account for ~31% of the total 1474 unique nouns recorded in the *Colin Charles Cree Dictionary* (195 *-win* nouns, 142 *-ikan* nouns, 57-*âkan* nouns, 64 *o-* nouns).

2.3.1.6 Lexical Prenouns

Woods Cree nouns may be modified by the addition of so-called lexical prenouns, prefixes which provide additional descriptive information about the noun in question. Semantically, these prenouns fill much the same function as adjectives in English:

121)	a.	*misi-sikâk*	b.	*maci-pisiskiw*
		misi-sikâk		maci-pisiskiw
		big(PN)-skunk(NA)		bad(PN)-animal(NA)
		'big skunk' (SM)		'bad animal, pest' (PK)
	c.	*mihkwasiniy iskwîw*	d.	*mitho-mîskanaw*
		mihko-asiniy iskwîw		mitho-mîskanaw
		red(PN)-stone(NA) woman(NA)		good(PN)-road(NI)
		'Red Stone Woman' (PN)		'the good road' (SM)

Prenouns may be applied recursively, although in the corpus, no more than two are ever seen on a single noun stem at once:

122)	a.	*kihci-mitho-mîcisowin* kihci-mitho-mîcisowin great(PN)-good(PN)-eating(NI) ‘being well nourished’ (PK)	b.	*kihci-thôski-tîhtapiwinihk* kihci-thôski-tîhtapiwin-ihk great(PN)-soft(PN)-chair(NI)-LOC ‘on the couch’ (SM)

Prenouns are often formed from nouns or verbs themselves through the addition of a final /i/, variously defined as a ‘particle-forming suffix’ (Wolvengrey 2022) or a linking vowel. The pronunciation of this /i/ is often almost imperceptibly subtle, although its presence may be inferred through its influence on stress:

123)	a.	*sîsîpi-sâkahikan* *sîsîp-i-sâkahikan* [duck-PFS/LV](PN)-lake(NI) ‘Duck Lake’ (PK)	b.	*pîwâpisko-cîmân* *pîwâpiskw-i-cîmân* [metal-PFS/LV](PN)-canoe(NI) ‘aluminium canoe’ (PK)

This suffixal *-i* is also used in compound nouns, where it is applied to the constituent acting as modifier. Since compound nouns in Woods Cree are (typically) left-branching, with the modifier preceding the head, this can lead to some ambiguity as to whether a complex noun is a compound or simply head noun with a productively fabricated lexical prenoun. Consider, for instance, example 123a; one might plausibly analyse this as a lexical prenoun (*sîsîpi-*, indicating a lake with some undefined relationship to ducks) or as a compound of *sîsîp* (‘duck’) and *sâkahikan* (‘lake’) with a linking vowel between them. In general, the exact semantics of individual prenouns can also differ depending on the referent which they are modifying:

124)	a.	*sîsîpi-pîway* sîsîp-i-pîway duck-PFS/LV-feather(NI) ‘duck feathers’ Note: Feathers from a duck	b.	*sîsîp-askihk* sîsîp-∅-ahkihk duck-PFS/LV-pot(NA) ‘kettle’ Note: A vessel shaped like a duck

Common prenoun-noun combinations may also become lexicalised, with varying degrees of compositionality:

125)	a.	*kisî-aya* kisî-aya venerable(PN)-being(NA) ‘elder’ (PK)	b.	*mistiko-nâpîw* mistiko-nâpîw wooden(PN)-man(NA) ‘carpenter’ (LR)[42]

[42] Compare semantics with *mistiko-cîmân* (‘wooden canoe’)

c. *mîtho-itithîhcikîwin*
mîtho-itithîhcikîwin
good(PN)-thought(NI)
'enjoyment' (UL)

d. *môsosiniy*
môso-asiniy
moose(PN)-stone(NA)
'bullet' (LR)

2.3.1.7 Structural Schematisation of the Woods Cree Noun

Based on the above observations of morpheme ordering in nouns, I have created the following template for nominal affixation in Woods Cree. According to this template, the possessive prefix is the leftmost possible affix, followed by lexical prenouns, of which multiple may occur (as indicated by the ⟨↺⟩ symbol. After the noun stem (which may be internally complex, such as in compounds or deverbal nominalisations), the closest suffix to the stem is the possessive theme suffix, followed by the diminutive suffix (although there is some variability in the relative ordering of these two elements, and the diminutive may occur twice). If both possessive theme and the diminutive suffix are present, the absentative goes between the two; right of the possessive theme suffix and left of the diminutive[43]. After these are the suffixes associated with plural possessors, followed by a slot which may contain the locative, distributive, vocative, number, or obviation suffixes; there is no relative ordering of these elements, as they are mutually exclusive.

Possessive Prefix	Lexical Prenoun (↺)	Noun Stem		Possessive Theme, Diminutive Suffix (↺)	Absentative	Diminutive Suffix (↺)	Plural Possessive Suffix	Locative, Distributive, Vocative, Number, Obviation
		(Verb Root)	*(Nominaliser)*					

Figure 6. Schematisation of the morpheme ordering in a theoretical maximally complex Woods Cree noun

This schematisation is sufficient to analyse the vast majority of inflected Woods Cree nouns; however, it is not without exception. One such exception may be found in the rare instances in which lexical prenouns are affixed to dependent nouns; in these cases, the prenoun occurs to the left of the possessive prefix:

126) *oski-wîwa*
oski-w-îw-a
new(PN)-3SG.POSS-wife(NDA)-OBV
'his new wife' (SM)

[43] The possessive theme and the absentative never co-occur in the corpus; however, my consultant (Solomon Ratt, personal correspondence via Arok Wolvengrey) consistently favoured the possessive theme preceding the absentative when both are present.

Unexpected morpheme orderings can also occur in nouns which have become lexicalised with an inflectional category suffix; consider the following example, where the diminutive occurs to the right of the distributive on a lexicalised distributive noun:

127) Highrock *omanicônâhkos*
Highrock o-manitôw-inâhk(w?)-is
Highrock 3SG.POSS-[spirit(NA)-DSTR](NI)-DIM
'Highrock's little spirit island' (PK)

However, leaving these exceptional cases aside, the expected relative ordering of the slots listed in Figure 6 may be demonstrated using the three following examples:

128) Possessive>Stem>Possessive Theme>Diminutive>Possessive Suffix>Number
kitawâsimisinawak
ki-t-awâsi-im-is-inaw-ak
2PL.POSS-LC-child(NA)-POSS.THM-DIM-2PL.POSS-PL
'our children' (PK)

129) Possessive>Lexical Prenoun>Stem>Possessive Theme>Possessive Suffix>Locative
nisîsîp-akohpiminâhk
ni-sîsîpi-akohp-im-inân-ihk
1PL.POSS-duck(PN)-blanket(NI)-POSS.THM.1PL.POSS-LOC
'our down-filled comforters' (SM)

130) Possessive>Stem>Absentative>Possessive Suffix>Number
kimosômipaninawak
ki-mosôm-ipan-inaw-ak
12PL.POSS-grandfather(NDA)-ABST-12PL.POSS-PL
'our (late) grandfathers' (SM)

2.3.2 Pronouns

Pronouns in Woods Cree may be broadly divided into four subcategories, personal, demonstrative, interrogative, and indefinite. Broadly, these pronouns must agree with the animacy, number, and obviation of their referent; however, in many instances, these paradigms appear to have simplified considerably from their historical predecessors.

2.3.2.1 Personal Pronouns

Woods Cree has two sets of personal pronouns, the simple and the additive-focal.

	Simple	Additive-Focal
1SG	*nîtha*	*nîsta*
2SG	*kîtha*	*kîsta*
3SG	*wîtha*	*wîsta*
1PL	*nîthanân*	*nîstanân*
12PL	*kîthânaw*	*kîstânaw*
2PL	*kîthawâw*	*kîstawâw*
3PL	*wîthawâw*	*wîstawâw*

Table 11. Personal pronoun paradigm

In pronouns, as elsewhere, Woods Cree distinguishes the inclusive first-person plural (12PL), which includes the listener, from the exclusive (1PL), which does not. However, unlike in other pronoun sets, there is no proximate-obviative distinction in personal pronouns. There is also no sex distinction in personal pronouns, or indeed anywhere in the inflectional landscape of Woods Cree; *wîtha* refers equally to males, females, and those of unknown sex.

Given the ubiquity of Woods Cree person-number agreement, personal pronouns are almost always optional, being primarily used to mark emphasis and contrastive focus. In particular, additive-focal personal pronouns are used to emphasise the inclusion of a given participant when that inclusion would not necessarily be presupposed based on preceding context; English translations typically render these as ‘I too’, ‘you too’, and so on.

131) a. *wîtha isa anihi â-kî-tôtamithit*

wîtha isa anihi (k)â-kî-tôta-mithit.
3SG reportedly PRON.DEM.MED.OBV CNJ-PST-do(VTI)-3`.CNJ
‘He was the one that did that.’ (PK)

b. *pît wîsta kî-wâpahtam*

pît wîsta kî-wâpaht-am
Pete 3SG.ADD.FOC PST-see(VTI)-3SG.IND
‘Pete saw it too’ (PK)
Context: Speaker was accused of lying about an improbable-sounding event.

2.3.2.2 Demonstratives

Woods Cree has a three-way distinction between proximate, medial, and distal demonstratives, with each set agreeing with animacy, number, and obviation of their referent. There is a degree of syncretism in this set, with the same demonstratives being used for inanimate plurals as are used for animate obviatives. Eastern varieties also have a distinct set of inanimate obviative demonstratives, used for inanimate nouns which are semantically obviative (that is, of lesser discursive relevance) but not necessarily morphologically marked with the obviative suffix:

132) *otinam manisam animîthiw wiyâs*

otin-am	manis-am	animîthiw	wiyâs
take(VTI)-3SG.IND	cut(VTI)-3SG.IND	DEM.MED.INAN.OBV	meat(NI)

'he took the meat and cut it' (PK)

Concerning demonstratives, all of the forms listed below in Table 12 are attested in the corpus for Eastern varieties, however, the only inanimate obviative demonstrative attested in the corpus for Western Woods Cree is *animîthiw*, which is found only in traditional stories:

	Proximate		Medial		Distal	
	Animate	Inanimate	Animate	Inanimate	Animate	Inanimate
3SG	*awa*	*ôma*	*ana*	*anima*	*nâha*	*nîma*
3PL	*ôko*, *ôki, ôhokok*	*ôho*	*aniki*(*k*)	*anihi*	*nîki*	*nîhi*
3`	*ôho*	*ômîthiw*	*anihi*	*animîthiw*	*nîhi*	*nîmîthiw*

Table 12. Demonstrative pronoun paradigm[44]

These demonstrative pronouns may occur both as noun modifiers and as verbal complements, with no associated change in form:

133) a. *awa kisîthiniw pî-sâkamaciwîw*

awa	kisîthiniw	pî-sâkamaciwî-w
DEM.PROX.ANIM.SG	old.man(NA)	come-climb.hill-3SG.IND

'this old man came walking in from up the hill' (PK)

b. *athisk ta-kî-mowât anihi â-kapatîsitâpît*

athisk	ta_kî-mow-ât	anihi	(k)â-kapatîsitâpî-t
because	would-eat(VTA)-3SG>3`.CNJ	DEM.MED.INAN.PL	CNJ-drag(VAI)-3SG.IND

'because it would have eaten them, the ones he dragged out.' (PK)

[44] The forms *anikik* and *ôhokok* are attested only at Pukatawagan

2.3.2.3 *kîkway* and *awîna*

The paradigms for *kîkway* (‘something’) and *awîna* (‘someone’) differ somewhat between Western and Eastern Woods Cree, with the Western paradigm being more syncretic than its Eastern counterpart. *kîkway* and *awîna* can also be used as interrogative pronouns, translating roughly to ‘what’ and ‘who’ respectively:

	kîkway		awîna
(3)SG	Western	*kîkway*	*awîna*(*k*)
	Eastern	*kîkway*(*i*), *kîkwan*	
(3)PL	Western	*kîkwaya*, *kîkwâtha*	*awîna*(*k*)
	Eastern	*kîkway*, *kîkwan*	*awina*, *awiniki*(*k*)
(3)`	Western	*kîkwâthiw*	*awîna*
	Eastern	*kîkwâthiw*, *kîkwan*	*awithiwa*, *awinaka*

Table 13. Pronominal paradigm for *kîkway* and *awîna*

As an interrogative pronoun, both varieties also have the form *kîko* as a morphosyntactically neutral alternative for *kîkway* and *awîna*, which remains formally static regardless of obviation or animacy.

134) a. *kîko misîl awa kâ-âcimat?*

kîko	misîl	awa	kâ-âcim-at
INTERR	Michel	DEM.PROX.ANIM.SG	CNJ-talk.about(VTA)2SG>3SG.CNJ

‘Which Michel are you talking about?’ (PK)

b. *kîko isa omaskisinisiwâwa?*

kîko	isa	o-maskisin-is-iwâw-a
INTERR	reportedly	3PL.POSS-moccasin-DIM-3PL.POSS-PL

‘Which are their little moccasins?’ (PK)

2.3.2.4 *kotak* and *iyako*

The paradigms for *kotak* (‘other’) and *iyako* (‘(that) same one’) are essentially identical between West and East, barring minor variance in pronunciation, the only major difference being the locative form of *kotak*, which is only attested in the West:

	Animate	Inanimate	Animate	Inanimate
3SG	*kotak*	*kotak*	*iyako*(*ni*), *iwako*	*iyako*
3PL	*kotakak*	*kotaka*	*iyakonik, iyakwanik*	*iyakoni*
3`	*kotaka*	*kotakîthiw*	*iyakoni, iyakwîthiw*	*iyakwîthiw*
LOC		*kotakihk*		

Table 14. Pronominal paradigm for *kotak* and *iyako*

2.3.2.5 *ayihîw* and *ayahâw*

ayihîw and *ayahâw*, described by Starks (1992:49) as 'whatever' and 'whoever' respectively, are a set of pronouns exclusive to Eastern Woods Cree; although in extremely frequent usage in Pukatawagan and South Indian Lake, I have found no attestation of them whatsoever in Stanley Mission or La Ronge, nor in the intermediate community of Pelican Narrows. *ayihîw* and *ayahâw* are both inflected to agree with number and obviation, with the inanimate *ayihîw* also having a locative form:

	Inanimate	Animate
3SG	*ayihîw*	*ayahâw*
3PL	*ayihîwa*	*ayahâwak*
3`	*ayihîthiw*	*ayahâwa*
LOC	*ayihîhk*	

Table 15. Pronominal paradigm for *ayihîw* and *ayahâw*

Semantically, these pronouns are used to refer to indefinite or generic referents, where the exact identity (or identities) of the referent(s) are of lesser relevance:

135) a. *ayahâwa isa ôho ithinito-cîmâna*
ayahâwa isa ôho ithinito-cîmân-a
whatever.ANIM.OBV reportedly DEM.PROX.ANIM.OBV old.style(PN)-canoe(NA)-OBV
'these traditional canoes' (PK)
Note: Discussing canoes in general, rather than any specific canoes

b. *kâ-kî-wâskânapicik mâna ayahâwak ithiniwak kâ-wîkihitocik*
kâ-kî-wâskânapi-cik mâna ayahâwak
CNJ-PST-sit.in.circle(VAI)-3PL.CNJ usually whatever.ANIM.PL
ithiniw-ak kâ-wîkihito-cik
person(NA)-PL CNJ-be.married(VAI)-3PL.CNJ
'they used to sit in a circle, those married people.' (PK)
Note: Discussing a historical practice on Treaty Days, when large numbers of married people would congregate to receive government rations

These pronouns in their various forms are also commonly used as filler words in Eastern Woods Cree:

136) a. *akwa ayihîw kâ-wî-âcimowân tânisîsa...*
akwa ayihîw kâ-wî-âcimo-wân tânisi isa
and whatever.INAN.SG CNJ-PRSP-tell.story-1SG.CNJ how reportedly
'Then, uh, I am going to tell how …' (PK)

2.3.2.6 *tân-* Interrogatives

According to previous literature (Starks 1992:56), the interrogative prefixes *tâna* ('which') and *tâniwâ* ('where') have obligatory number, animacy, and obviation agreement in the same manner as the previous pronouns discussed. However, there is a substantial discrepancy between the richness of these paradigms as described in the literature and the observed forms of these pronouns in the corpora, with most reported forms being entirely unattested in Western or Eastern varieties. Like *kîkway* and *awîna*, forms of *tâna* and *tâniwâ* can also be used as non-interrogative pronouns, with the indefinite meanings 'that one' and 'there' respectively:

	tâna		*tâniwâ*	
	Animate	Inanimate	Animate	Inanimate
3SG	*tâna*, *tânitowihk*, *tânîtwâ*[45]	*tânima*, *tân*(*i*), *tânitowihkân*	*tâniwâ*	*tâniwî*
3PL	*tânihkâ*	*tânîhi*, *tânitowihkân*	*tân*(*iw*)*îhkâ*	*tân*(*iw*)*îhâ*
3`	*tâniwîhi*		*tân*(*iw*)*îhâ*	

Table 16. Pronominal paradigm for *tâna* and *tâniwâ*

Not only do these *tân-* interrogative paradigms appear to be reducing in morphological complexity across dialects, they also appear to be being gradually displaced by non-inflecting

[45] *tânîtwâ* is attested in Pelican Narrows.

interrogative pronouns with similar semantics. *tâniwâ* in particular seems almost completely replaced by *tânitî* and *tânita*, both of which are non-inflecting:

137) *tânitî îtokwî wîtha owîcîwakana?*
tânitî îtokwî wîtha o-wîcîwakan-a
where DUB FOC 3SG.POSS-friend(NA)-OBV
'Where were his friends?' (SM)

In addition to *tâna* and *tâniwâ*, a number of other (non-inflecting) *tân-* interrogatives also exist in Woods Cree, formed by combining the *tân-* prefix with a particle (see Section 2.3.3.1) indicating manner or degree. These include:

138) a. *tânithikohk*
tân-ithikohk
[INTERR-to.such.an.extent]
'how much, to such an extent' (LR)

b. *tânitahto*
tân-itahto
[INTERR-so.many]
'how many' (LR, SM)

2.3.3 Particles

The label of 'particle' in Woods Cree does not so much constitute an internally coherent morphosyntactic class of its own as much as it does a catch-all term for all non-inflecting lexical items which cannot be classed as nouns, pronouns, or verbs. In layman's terms, I have heard it referred to by Algonquianists, quite aptly, as the '"whatever" class' of Cree words (Katherine Schmirler, personal correspondence). Particles in Woods Cree largely fulfil the same functions as particles in other Western Cree dialects (Wolvengrey 2022); however, where previous authors have classed lexical prenouns (see Section 2.3.1.6) and preverbs (see Section 2.3.4.6) as particles[46], I will restrict my definition to include only those lexical items which may occur as independent units. I divide this set into two broad subcategories, that of Indeclinable Particles (or IPCs, in Algonquianist terminology) and that of Interjections.

2.3.3.1 Indeclinable Particles

Indeclinable particles are extremely diverse in function, and are broadly unified only through their exclusion from other word classes and their general lack of inflection. Some of their most common functions include:

139) Adpositions or location markers

otâhk	*wathawîtimihk*	*ispimihk*
'behind'	'outside'	'above'

[46] Indeed, in syllabics, they are often written as space-separated units.

140) Temporal markers

kayâs	*anohcihkî*	*acithaw*
‘long ago’	‘recently’	‘temporarily’

141) Adverb-like descriptors

nisihkâc	*nâspic*	*kwayask*
‘slowly’	‘greatly’	‘properly’

142) Quantifiers

âtiht	*kahkithaw*	*tahto*
‘some’	‘all’	‘every’

143) Negation

mwâc	*namôtha*	*îkâ*
‘no, not’	‘no, not’	‘no, not’

144) Conjunctions

mâka	*athisk*	*ikwa*
‘but’	‘because’	‘and’

145) Evidentials

isa	*kîhcinâc*	*îtokwî*
‘reportedly’	‘surely’	‘probably’

146) Question markers

cî
‘Closed Interrogative Marker’

Although generally non-inflecting, IPCs are not entirely immune to certain morphological processes. For instance, systematic diminutives exist for many particles, formed using largely the same processes as are used in noun diminutivisation (although the extent of the productivity of this phenomenon remains to be seen through fieldwork):

147)	a.	*kinwîsk*	>	*kinwîskîs*
		‘a long time’		‘a little while’
148)	a.	*mistahi*	>	*miscahîs*
		‘lots’		‘a little extra’

Similar to verbs (see Section 2.3.4.5), IPCs can also be reduplicated according to one of two patterns. The first of these follows the pattern *C1a-* (or *ay-* for vowel-initial stems), and is somewhat obscure in its semantic function; increased emphasis seems to be engendered in most cases, but in some instances, it seemingly has no effect at all. The second reduplicative pattern is

C1âh- (or *âh-*[47] for vowel-initial stems); this adds either a distributive meaning or discontinuous temporally iterative meaning, depending on the nature of the particle:

149)	a.	*âsay* 'already'	>	b.	*ay-âsay* 'all of the sudden'
	c.	*âpihtaw* 'half'	>	d.	*ay-âpihtaw niska* 'hybrid goose'[48] (SM)
	e.	*piyâhtik* 'carefully, slowly'	>	f.	*pa-piyâhtik* 'carefully, slowly'
	g.	*acithaw* 'temporarily'	>	h	*âh-acithaw* 'in short spurts' (SM)
	i.	*pîyak* 'one'	>	j.	*pâh-pîyak* 'for each one, one each'

In addition to these, there exist many set combinations of IPCs which form lexicalised phrases, the meanings of which are often partially or entirely non-compositional. These inseparable 'particle phrases' often assume syntactic distributions akin to particles themselves:

150)	a.	*cîskwa pitamâ* cîskwa pitamâ yet temporarily 'Wait a minute'	b.	*matwân cî* matwân cî is.it.really INTERR 'I wonder, I don't know'

Finally, Cree numerals are typically classed as indeclinable particles or particle phrases, and I will likewise class them as such here; however, they are sufficiently distinct to warrant further discussion. I provide a description of the Woods Cree numeral system in Appendix F.

2.3.3.2 Interjections

Interjections differ from IPCs in two key respects; firstly, they consistently constitute syntactic holophrases, requiring no surrounding lexical context or integration into a larger phrase structure, and secondly, they are often semantically ideophonic, serving to illustrate or augment an emotional or physical state. (see Wolvengrey 2013).

[47] Or, more rarely, *ây-*
[48] Used in a text describing species of birds, specifically to contrast with non-hybrid geese

151) a. *wahwâ! kwayask nikî-mithwâpisinin*
wahwâ kwayask ni-kî-mithwâpisin-in
INTERJ properly 1SG-IND-PST-see.pleasing.sight-1SG-IND
'Whoa! I liked what I saw.' (SM)

b. *toni napakatihkwîwak, pôy!*
(mi)toni napakatihkwî-wak pôy
very be.flat.faced-3PL-IND INTERJ
'Boy! They really had flat faces' (PK)

2.3.4 Verbs

The languages of the Cree dialect continuum have often been described as 'verb-based', owing to their reliance on verbs to serve semantic functions often accomplished in other languages by adjectives, adverbs, or even nouns, as well as due to the considerable morphological complexity of these verbs. Woods Cree is no exception to this trend; in the *Colin Charles Cree Dictionary*, for example, 3539 of the total 5858 distinct lexical entries are verbs, compared to 1474 nouns. In this section, I will first discuss the internal structure of Woods Cree verb stems before describing the various inflectional categories for which these stems may or must receive affixation, providing paradigms where necessary. Finally, I will also provide a full schematisation for a theoretically maximally complex verb, providing corpus evidence to corroborate this schematisation throughout.

2.3.4.1 Internal Structure

Internally, Woods Cree verb stems must contain at least two obligatory elements, the Initial (or Root) and the Final (of which there may be several). Stems may also have an optional Medial element intervening between the Initial and Final(s). Generally, the Initial in a verb stem conveys the basic semantic sense of the verb in question, with Medials increasing the specificity of this meaning (often manifesting in the form of incorporated nouns), and Finals indicating the nature or the means of occurrence of the action or state being described:

152) Initials[49]

wâp-	*nip-*	*ot*
'clear, bright, white'	'die, death, sleep'	'pull'

153) Finals

-am	*-i*	*-ah*	*-nakw*
'by eye, by vision'	'to be'	'by external medium'	'appear thus'

[49] Morpheme glosses sourced from Cook & Muehlbauer (2010)

154) Medials

Classificatory Medials

-nisk	*-askw*	*-âpisk*
‘hand, arm’	‘wood’	‘stone, metal’

Incorporated nouns

-iskwîw	*-stikwân*	*-kinosîw*
‘woman’	‘head’	‘fish’

155) Initial + Final

a. *wâpam-*
wâp-am
[clear(INIT)-by.eye(FIN)](VTA)
‘to see s.o.’

b. *nipi-*
nip-i
[die(INIT)-to.be(FIN)](VAI)
‘to be dead’

156) Initial + Final + Final

wâpinâkosi-
wâp-inâkw-isi
[white(INIT)-appear.thus(MED)-stative(FIN)](VAI)
‘to look white’

157) Initial + Medial + Final

otâskwah-
ot-âskw-ah
[pull(INIT)-wood(MED)-by.external.medium(FIN)](VTI)
‘to pull s.t. using a stick’

158) Initial + Medial + Final+ Final

masinâpiskahikâso-
masin-âpisk-ah-ikâso
[mark(INIT)-stone(MED)-by.external.medium(FIN)-reflexive(FIN)](VAI)
‘paint [an image of] oneself on stone’

This general pattern of verbal stem structure is shared, with some variation, among all Algonquian languages (Goddard 1990). Verbal inflection in Woods Cree occurs entirely through prefixation and suffixation, with stems themselves remaining unchanged (aside from in morphophonological processes occurring at affix boundaries). This is in contrast to other Western Cree dialects, particularly historically, whereby vowel alternations (called ‘Initial Change’) may be applied to the stem to indicate shifts in mode (see 2.3.4.2.3 Subjunctive).

Most obligatory verbal inflection occurs through suffixation, thus interacting with the Final, as do many derivational suffixes such as nominalisers (see Section 2.3.1.5). The Final is also the

element of the verb stem which determines its verb class (see Section 2.3.4.7), indicating the verb's transitivity and the potential animacy of its referents .

159) a. *wîpinikâtîw*
[wîp-in-ikâtî]-w
[discard(INIT)-do.by.hand(FIN)-middle.reflexive(FIN)](VII)-3SG.IND
'it is thrown away' (LR)

b. *wîpinikîw*
[wîp-in-ikî]-w
[discard(INIT)-do.by.hand(FIN)-general.goal(FIN)](VAI)-3SG.IND
's/he throws (things) away' (LR)

c. *wîpinam*
[wîp-in]-am
[discard(INIT)-do.by.hand(FIN)](VTI)-3SG.IND
's/he throws it away, s/he gets rid of it' (LR)

d. *wîpinîw*
[wîp-in]-îw
[discard(INIT)-do.by.hand(FIN)](VTA)-3SG>3`.IND
's/he throws him/her away, s/he gets rid of him/her' (LR)

In stems with multiple Finals, it is always the rightmost Final which determines the class of the verb. Consider the following example; the *-ihkî* Final is used to form VAI stems, but when the secondary Final *-aw* is added, the stem becomes a VTA instead:

160) *âhkosîwikamikohkaw-*
âhkosîwikamikw-ihkî-aw
[hospital(INIT)-create(FIN)-benefactive(FIN)](VTA)
'to build a hospital for s.o.' (PK)

A wide variety of units may act as Initials; nouns, even those derived from deverbal nominalisation, may be converted directly into Initials to create novel verb stems:

161) *wâskahikanihkîw*
[wâskahikan-ihkî]-w
[house(INIT)-create(FIN)](VAI)-3SG.IND
's/he builds houses' (LR)
Note: *wâskahikan* is itself a nominalised form of the verb stem *wâskah-* ('to encircle')

Given their internal modularity, the spontaneous creation of novel verb stems by fluent speakers is commonplace, often as a point of humour. Consider the following example, where a fluent speaker incorporates the English word ‘roll’ as a Medial in a playful turn of phrase:

162) *î-papâmi**rol**îwit*
î-[papâmi-**rol**-iwi]-t
CNJ-[around(INIT)-**roll**(MED)-to.be(FIN)](VAI)-3SG.CNJ
‘to roll around’ (PN)
Note: Compare to expected form *papâmitihtipî-*

2.3.4.2 Order

As with other Algonquian languages, all Woods Cree verbs have several parallel sets of distinct inflectional affixes used to encode person and number, with different sets used in different syntactic contexts. These sets are known in Algonquianist terminology as Orders, with Woods Cree having three such Orders. These are the Independent, used on verbs which head matrix clauses, the Conjunct, used (mostly) for verbs heading subordinate clauses, and the (self-explanatory) Imperative:

163) a. *kitapin*
ki-t-api-n
2SG.IND-LC-sit(VAI)-2SG.IND
‘you sit’ (LR, PK)

b. (*niwâpahtîn*) *î-apiyin*
ni-wâpaht-în î-api-yin
1SG.IND-see(VTI)-1SG.IND CNJ-sit(VAI)-2SG.CNJ
‘(I see that) you sit’ (KS)

c. *api*
api-∅
sit(VAI)-2SG.IMP.IMM
‘Sit!’ (SM, LR)

Historically, the Woods Cree Order system was considerably more complex (see, for instance, Bloomfield 1946). The Independent Order was subdivided into two Modes, Neutral (used for general purpose indicatives) and Dubitative (used for uncertain or unverifiable events, as well as some interrogatives), while the Conjunct was divided into three such Modes, Neutral, Dubitative, and Subjunctive (used chiefly for conditional events). These Modes were further subdivided into tense variants, with different affix sets for Neutral (or Present) tense and Preterit tense for Modes in the Independent Order and Conjunct Order, and Immediate and Delayed tense variants for the Imperative Order. These tense variants were then further subdivided into Relational forms (which implied the participation of actors in the verb which were not specifically mentioned) and Non-Relational forms (which did not).

By contrast, modern Woods Cree has entirely lost the Neutral-Dubitative, Neutral-Preterit, and Relational-Non-Relational distinction (see Appendix G). Only the Neutral-Subjunctive Mode distinction in the Conjunct and the Delayed-Immediate Tense distinction in the Imperative remain to attest to the historical intricacy of this system. As such, counting only fully productive paradigms, the modern Woods Cree Order system may be schematised as follows:

Order	Mode	Tense
Independent	Neutral	Neutral
Conjunct	Neutral	Neutral
	Subjunctive	Neutral
Imperative		Immediate
		Delayed

Table 17. The contemporary Woods Cree verbal Order system

This simplified Order system is essentially identical to the equivalent system in Plains Cree (Wolfart 1973:41); this is in contrast to Swampy Cree and Moose Cree, which have both retained the more complex historical system (Ellis 2016:136)

2.3.4.2.1 Independent and Conjunct

The Independent and Conjunct overlap slightly in function, but differ significantly in form. Independent affixes encode person and number through coordinated sets of prefixes and suffixes. Conjunct affixes, by contrast, only encode person and number through suffixes, and use an entirely unrelated set of suffixes for this purpose to the Independent. Generally, Conjunct verbs must also bear one of several possible Conjunct prefixes. For example:

164) a. *ninipahison* ≈ *î-nipahisoyân*
ni-nipahiso-n — î-nipahiso-yân
1SG.IND-kill.self(VAI)-1SG.IND — CNJ-kill.self(VAI)-1SG.CNJ
‘I kill myself’ — ‘I kill myself’

b. *nipahisow* ≈ *î-nipahisot*
nipahiso-w — î-nipahiso-t
kill.self(VAI)-3SG.IND — CNJ-kill.self(VAI)-3SG.CNJ
‘s/he kills him/herself’ — ‘s/he kills him/herself’

Both Independent and Conjunct Order can be used for basic declaratives and interrogatives, and both Orders may occur on verbs heading matrix clauses. However, verbs in subordinate clauses must always take Conjunct morphology:

165) Independent Verb in Matrix Clause
mihcîtiwak ikotî sâkwîsiwak
mihcîti-wak ikotî sâkwîsiw-ak
be.many(VAI)-3PL.IND there mink(NA)-PL
'Over there, there are many mink' (SM)

166) Conjunct Verb in Subordinate Clause
nikiskîthihtîn ita î-mihcîticik kiyâskwak
ni-kiskîthiht-în ita î-mihcîti-cik kiyâskw-ak
1SG.IND-know(VTI)-1SG.IND place CNJ-be.many(VAI)-3PL.CNJ gull(NA)-PL
'I know of a place where the gulls are so numerous' (SM)

167) Conjunct Verb in Matrix Clause
kâhkâkiwak aniki kâ-mâwaci-mihcîticik pithîsîsak
kâhkâkiw-ak aniki kâ-mâwaci-mihcîti-cik pithîsîs-ak
raven(NA)-PL DEM.PROX.MED.PL CNJ-superlative-be.many(VAI)-3PL.CNJ bird(NA)-PL
'Common Ravens are some of the world's most cosmopolitan birds.' (SM)

168) Conjunct Verb in Interrogative
â-kî-ati-mihcîtiyâ pîthisk anohc, ciyi?
(k)â-kî-ati-mihcîti-yâ(hk) pîthisk anohc ciyi
CNJ-PST-start-be.many(VAI)-1PL.CNJ finally today INTERR
'There were a lot of us today, right?' (PK)

Unlike the Independent, the Conjunct is tense neutral by default, and can be used without tense prefixes to describe future or past events where an Independent verb would require explicit tense marking (although tense marking may still be optionally used). Consider the following example, where the verb *ayâ-*, which is conjugated in the Independent, must have the past-tense marker *kî-*, whereas *itohtî-* and *kiyokî-*, which are in the Conjunct, do not:

169) *îkotî kî-ayâwak, îkotî îtokî î-itohtîcik î-kiyokîcik*
îkotî kî-ayâ-wak îkotî îtokî î-itohtî-cik
there PST-be.there(VAI)-3PL.IND there DUB CNJ-go(VAI)-3PL.CNJ
î-kiyokî-cik
CNJ-visit(VAI)-3PL.CNJ
'They were there, perhaps going over there for a visit' (PK)

When used in matrix clauses instead of the Independent, the Conjunct conveys discursive uncertainty or unboundedness, and is frequently used to describe hypothetical events, or to relay secondhand information for which the speaker has no direct evidence. Thus, in matrix clauses, the Conjunct appears to have absorbed some of the functions of the now-disused Dubitative paradigm:

170) a. *pî-wâpan nitisithihkâson*
pî-wâpan ni-t-isithihkâso-n
pî-wâpan 1SG.IND-LC-be.named(VAI)-1SG.IND
'My name is Pî-Wâpan' (PK)

b. *tânisi îtokî â-isithihkâsocik*
tânisi îtokî (k)â-isithihkâso-cik
how DUB CNJ-be.named(VAI)-3PL.CNJ
'I don't know what they call themselves' (PK)

2.3.4.2.2 Conjunct Prefixes

In addition to suffixes indicating person-number-obviation agreement, Neutral Conjunct verbs must also have one of either one of two possible Conjunct prefixes *î-* or *kâ-*, a tense prefix, or a modality prefix. The distinction between *î-* and *kâ-* is subtle, and primarily interfaces with discourse structure. *î-* is generally used to signal new, discursively important information, such as when establishing a new discursive frame of reference or implying continuity between multiple such frames of reference. *kâ-*, by contrast, is used to describe events occurring within an already established frame of reference which do not significantly alter this frame (Starks 1992:222-47). Consider the following passage; *itohtîho-* ('to travel') is affixed with *î-* because it relays new information which establishes a new frame of reference in the discourse, whereas *sipwîhtî-* ('to leave') is affixed with *(k)â-* because it describes an event within a previous established frame of reference.

171) *kâ-âcimostâkowâ mâna* five thousand miles *î-kî-itohtîhot atimwa ohci. pâhkaci â-sipwîhtît ôma akwâni misiwî*
kâ-âcimostaw-ikowâ(hk) mâna five thousand miles
CNJ-tell.story(VTA)-3SG>1PL.CNJ usually five thousand miles

î-kî-itohtîho-t atimw-a ohci pâhkaci
CNJ-PST-travel(VAI)-3SG.CNJ dog(NA)-OBV with at.that.time

(k)â-sipwîhtî-t ôma akwâni misiwî
CNJ-depart(VAI)-3SG.CNJ FOC then everywhere
'He used to tell us he had travelled five thousand miles by dog team ('with dogs'). Once he travelled out, he would go all over' (PK)

The exact boundaries of these discursive distributions are highly malleable, and it is difficult to predict with certainty whether *î-* or *kâ-* will occur on a given Conjunct verb. This is further complicated by the fact that *kâ-* is also used as a relativiser regardless of its position in the discursive frame of reference:

172) *pîthisk ana kâ-nîhithawinâkosit ayamihikimâw kâ-tîpwâtiht ta-sâh-sihkimâwasot.*

pîthisk	ana	kâ-nîhithawinâkosi-t	ayamihikimâw
finally	DEM.MED.ANIM.SG	CNJ-look.Cree(VAI)-3SG.CNJ	priest

kâ-tîpwât-iht	ta-sâh-sihkimâwaso-t
CNJ-call(VTA)-X>3SG.CONJ	FUT-RDPLS-encourage(VAI)-3SG.CNJ

'Eventually the one who looked like a Cree was called up to give encouragement.' (SM)

2.3.4.2.3 Subjunctive

The Subjunctive (or Future-Conditional) is used to express conditional mood (typically glossed in English as 'if' or 'when'), and often co-occurs with IPCs to disambiguate the nature and certainty of the conditionality which it expresses. Although I classed it in Section 2.3.4.2 as a sub-Mode of the Conjunct due to its historical origins as such, from this point on I will treat it as its own distinct Order-level category, as this analysis more closely resembles its behaviour in the modern language. The Woods Cree Subjunctive is relatively simple compared to its neighbours. In Swampy Cree and some dialects of Plains Cree, the Subjunctive is subdivided into the Future Conditional (e.g. *apiyahki* ('when we sit') (Wolvengrey 2022)) and the Timeless Conditional (e.g. *êpiyahki* ('whenever we sit') (Wolvengrey 2022)). By contrast, in Woods Cree, only the Future Conditional form exists.

173) a. *kiskîthîtamowin nitawîthihtamani*

kiskîthîtamowin	nitawîthiht-amani
information(NI)	need(VTI)-1SG.SUBJ

'if you need information' (UL)

b. *ispî takosiniyahki kika-kotawân.*

ispî	takosin-iyahki	ki-ka-kotaw-ân
when	arrive(VAI)-12PL.SUBJ	2SG.IND-FUT-build.fire-2SG.IND

'When we arrive you will make a fire.' (SM)

The Subjunctive uses a virtually identical set of suffixes to the (Neutral) Conjunct, modified by the addition of an /i/ to the rightmost edge[50]. If the relevant Conjunct person-number suffix ends

[50] On the 2PL suffix, which has an underlying final /w/, the resultant /wi/ sequence becomes /o/

in a /t/, this /t/ affricates to /t͡s~t͡ʃ/ . For a full demonstration of the formal similarities between the Subjunctive and Conjunct, see Table 20 in Section 2.3.4.7.2.

174)	a.	*î-nipiyân*	>	*nipiyâni*
		î-nipi-yân		nipi-yâni
		CNJ-die(VAI)-1SG.CNJ		die(VAI)-1SG.SUBJ
		'I die' (NC)		'when I die' (NC)
	b.	*î-nipit*	>	*nipici*
		î-nipi-t		nipi-ci
		CNJ-die(VAI)-3SG.CNJ		die(VAI)-3SG.SUBJ
		's/he dies' (NC)		'when s/he dies' (NC)

Like Conjunct verbs, Subjunctive verbs do not require tense marking, and given their often irrealis semantic function, are typically seen without it. However, tense-marking on Subjunctive forms is attested in the corpus:

175) *kî-nipahiyani pîyakwan îkosi ta-tôtawin*

kî-nipah-iyani	pîyakwan	îkosi	ta-tôtaw-in
PST-(VTA)-2SG>1SG.SUBJ	same	thus	RDPLW-do(VTA)-2SG>1SG.IMP.IMM

'After you kill me, you can do the same with me' (PK, modified[51])

2.3.4.2.4 Imperative

The Imperative Order is used for direct commands or exhortations, and correspondingly only inflects for first-person plural inclusive and second-person subjects. Two sets of Imperative suffixes exist, the Immediate, used for issuing commands to be followed immediately, and the Delayed, used to issue prospective commands to be followed later. As mentioned, the Imperative is the only Order which has retained multiple suffix sets to indicate tense, rather than using (unrelated) tense prefixes:

176)	a.	*itohtî*	>	*itohtîhkan*
		itohtî-∅		itohtî-ihkan
		flee(VAI)-2SG.IMP.IMM		flee(VAI)-2SG.IMP.DEL
		'Go!'		'Go later!'
	b.	*itohtîk*	>	*itohtîhkîk*
		itohtî-k		itohtî-ihkîk
		flee(VAI)-2PL.IMP.IMM		flee(VAI)-2PL.IMP.DEL
		'(Everybody) go!'		'(Everybody) go later!'

[51] Several embedded clauses were removed to reduce the length of this sentence, which may be found in ME-13

c. *itohtîtân* > *itohtîhkahk*
itohtî-itân itohtî-ihkahk
flee(VAI)-12PL.IMP.IMM flee(VAI)-12PL.IMP.DEL
'Let's go!' 'Let's go later!'

Although the Delayed Imperative is fully productive, the Immediate form is by far the more common and less marked of the two constructions. Often, the semantic function of the Delayed Imperative may be accomplished through Independent Order verbs with future tense markers:

177) *na-nitonawâw*
na-nitonaw-âw
1SG.FUT-seek(VTA)-1SG>3SG.IND
'Let's go look for him!' (PK)

The distinction between the Immediate and Delayed Imperative is slightly more complex than a simple tense distinction. The Immediate Imperative, for example, is almost always used for commands which lack a specific timeframe. For instance:

178) a. *ikâwîtha kocihtâ ôma kîkihk*
ikâwîtha kocihtâ-∅ ôma k-îki-ihk
NEG try(VTI)-2SG.IMP.IMM FOC 2SG.POSS-home(NDI)-LOC
'Don't try this at home! (in general)' (SM)

b. *ispî mâna kâ-kakwâtaki-kwîtawîthimat awiyak sâkihiso nawac sôhki*
ispî mâna kâ-kakwâtaki-kwîtawîthim-at awiyak
when usually CNJ-dreadfully-miss(VTA)-2SG>3SG.CNJ someone
sâkihiso-∅ nawac sôhki
love.self-2SG.IMP.IMM more strongly
'On those days when you miss someone the most, love yourself harder.' (SM)

Not only this, but the Immediate Imperative can also be used instead of the Delayed Imperative for certain delayed commands, that is, commands which cannot be accomplished immediately (see example 179a). In these instances, it can function as a point of emphasis, indicating that a command should be accomplished as soon as possible, even if not 'immediately' in the strictest sense. The Delayed Imperative, by contrast, is generally used for prospective commands which are to be accomplished at an indeterminate time in the future, and seldom imply great urgency (see example 179b):

179) a. *hâw waniskâk! kîmâsik kâtha mistahi itwînamok! pônamok, kisâkamisikîk.*

hâw	waniskâ-k	kîmâsik	kâtha	mistahi
okay	wake.up(VTI)-2PL.IMP.IMM	quietly	NEG	very

itwînamo-k pônamo-k
make.noise(VAI)-2PL.IMP.IMM build.fire(VAI)-2PL.IMP.IMM

kisâkamisikî-k
make.tea(VAI)-2PL.IMP.IMM
'Okay, wake up! Don't make too much noise! Make a fire, make tea!' (PK)

b. *costân kîtahtawîth nitawi-wâpahtamohkan câh-cikâstîpitamôhkan.*

costân	kîtahtawîth	nitawi-wâpahtamo-ihkan
just.in	then	go-take.a.look(VAI)-2SG.IMP.DEL

câh-cikâstîpitamo-ihkan
RDPLS-take.photograph(VAI)-2SG.IMP.DEL
'Just on the spur of the moment, go and take a look, take pictures of it.' (PK)

2.3.4.3 Tense

Aside from in the aforementioned Imperatives, tense in modern Woods Cree is communicated entirely through prefixes. For Independent verbs, past tense is almost always indicated through the prefix *kî-*. For Conjunct verbs, past tense may be optionally indicated using the same *kî-* suffix. Starks also notes that *kî-* may also be used on Conjunct verbs without the normal Conjunct prefix(es) (i.e. *î-* or *kâ-*) to indicate irreality:

180) a. *âsay kî-ihtakwan îkospî* lard

âsay	kî-ihtakwan	îkospî	lard
already	PST-exist(VII)-3SG	at.that.time	lard

'Lard was already there' (PK)

b. *mwâc itha kîkway sâsîskîkwân îkamâ kî-ohci-ihtakwa pîwâpisk*

mwâc	itha	kîkway	sâsîskîkwân	îkamâ	kî-ohci-ihtakwahk
NEG	FOC	PRON.INAN.SG	frying.pan(NI)	none	PST-from-exist(VII)-3SG.CNJ

pîwâpisk
iron(NI)
'There was no frying pan at all, obviously [because] iron did not exist' (PK)

An additional past tense prefix, *ôh-*, is also attested throughout the corpus; however, it is substantially less common than *kî-* (34 instances of *ôh-* compared to 4445 instances of *kî-*). *ôh-* is used almost exclusively in negative clauses, and strongly prefers Independent first and second-person subjects.

181) a. *môtha n-ôh-misikitin*
môtha ni-ôh-misikiti-n
NEG 1SG.IND-PST-be.big(VAI)-1SG.IND
'I was not old (then)' (PK)

b. *pô mîna môtha n-ôh-nisitohtawâwak atîht ithiniwak*
poko mîna môtha ni-ôh-nisitohtaw-âwak
only also NEG 1SG>3PL.IND-PST-understand(VTA)-1SG>3PL.IND
atîht ithiniw-ak
some people(NA)-PL
'And I also really didn't understand some people' (SL)

Present tense is unmarked on Independent, Conjunct, and Subjunctive verbs, although partial reduplication may be used to indicate progressive aspect (see Section 2.3.4.5).

The prefixes used for marking future tense differ between Western and Eastern varieties. Western Woods Cree uses three prefixes, *ka-*, *ta-*, and *kita-*. The distributions of these three prefixes are mutually overlapping, and there is no apparent semantic distinction between them[52]. However, broadly speaking, *ka-* is preferred for first-person and second-person Independent verbs, while *ta-* is preferred for third-person Independent forms and all Conjunct forms. *kita-* also usually occurs on Conjunct verbs, but seems to have no person or number preference. On various occasions, multiple different future tense markers are seen co-occurring on the same verb stem in the corpus (see example 182). The function of this is unclear; my consultant suspects that they may be speech errors (Solomon Ratt, personal communication via Arok Wolvengrey):

182) *môtha awinak ta-ka-otihtikonaw*
môtha awinak ta-ka-otiht-ikonaw
NEG PRON.ANIM.SG FUT-FUT-reach(VTA)-3SG>12PL.IND
'Nobody will get us (harm us)' (PK)

Through haplology, when the *ka-* prefix is combined with the second-person Independent prefix *ki-*, the resultant form *kika-* is frequently shortened to *ka-* in both Western and Eastern varieties:

183) *matwân cî ka-kaskihtân*
matwân cî (ki)-ka-kaskihtâ-n
I.wonder (2SG.IND)-FUT-be.able(VTI)-2SG.IND
'I wonder if you can' (SM)

[52] Perhaps unsurprisingly, as all three descend from the PrA future marker **kata-*

Eastern Woods Cree has an additional future marker, *na-*, a portmanteau used to mark future tense for first-person Independent verbs:

184) *môtha itha kîhtwâm na-itôhikân.*
môtha itha kîhtwâm na-itôhik-ân
NEG FOC again 1SG.FUT-point(VAI)-1SG.IND
'I am not ever going to point again.' (PK)

The distributions of *ka-* and *ta-* are similar between East and West, however, *kita-* is much less common in both. Greensmith (1985b:67) claims that *ta-* and *kita-* only occur on Conjunct verbs in Eastern dialects; however, I have found corpus examples with verbs from all four verb classes which contradict this. For instance:

185) *kita-kitimâkan askiy*
kita-kitimâkan-∅ askiy
FUT-be.pitiful(VII)-3SG.IND land(NI)
'the land will be in a sorry state' (PK)

ka- and *ta-* can also combine with the past tense prefix *kî-*, however, the resultant semantics are non-compositional, and convey information related to modality. *ta-kî-* is usually translated as 'should', implying expectation or recommendation, and *ka-kî-* is usually translated as 'can', implying ability or permission. *ta-kî-* strongly prefers Conjunct verbs (although it can be affixed to Independent verbs), whereas *ka-kî-* does not seem to have any strong Order preference:

186) a. *ta-kî-âmaciwîyan anihi waciya*
ta_kî-âmaciwî-yan anihi waciy-a
should-climb(VAI)-2SG.CNJ DEM.MED.INAN.PL mountain(NI)-PL
'you were supposed to climb those mountains' (SM)

b. *mwâc kapî ka-kî-pa-pamihitinâwâw*
mwâc kapî (ki)-ka_kî-pa-pamih-itinâwâw
NEG always (1SG>2PL.IND)-can-RDPLW-look.after(VTA)-1SG>2PL.IND
'I will not always be able to look after you' (PK)

Given their divergence in function, I will analyse *ta-kî-* and *ka-kî-* as distinct, non-compositional modal prefixes in this investigation (see Section 2.3.4.4) rather than as simple combinations of tense prefixes; however, the precise relationship between *ta-kî-*, *ka-kî-*, and their constituent elements is a matter well-deserving of further study.

Although generally tense-neutral, Conjunct verbs may eschew their typical Conjunct prefix (i.e. *î-* or *kâ-*) to take a (future) tense prefix in its stead. However, in these instances, even bare tense markers often take up a modal, rather than purely temporal, function:

187) *îka ta-kosta awinak ta-atoskît*
îka ta-kost-a(hk) awinak ta-atoskî-t
NEG FUT-fear(VTI)-3SG.CNJ PRON.ANIM.SG FUT-work(VAI)-3SG.CNJ
'nobody would be afraid to work' (PK)

In addition to *na-*, *ka-*, *ta-*, and *kita-*, both varieties also frequently use the prefix *wî-* to indicate future tense. However, the *wî-* prefix is notably distinct from the other future prefixes in its implication of intentionality. Starks describes *wî-* as occurring in "clauses which refer to a desire", with Greensmith similarly defining the prefix as 'intend to, want to':

188) a. *î-wî-môwit ana*
î-wî-môw-it ana
CNJ-PRSP-eat(VTA)-3SG>1SG.CNJ DEM.MED.ANIM.SG
'He's gonna eat me.' (SM)

b. *kâ-wî-mîciyâ kîkwan*
kâ-wî-mîci-yâ(hk) kîkwan
CNJ-PRSP-eat(VTI)-1PL.CNJ PRON.INAN.SG
'when we wanted to eat something' (PK)

wî- is also combinatorially distinct from the other future tense markers; unlike *na-*, *ka-*, *ta*, and *kita-*, *wî-* generally co-occurs with the Conjunct marker(s) *î-* and *kâ-* when used on Conjunct Order verbs, rather than replacing them. Additionally, *wî-* can also combine with future tense markers such as *ta-* to form semantically compositional units (unlike the non-compositional tense-tense modals *ta-kî-* and *ka-kî-*)

189) *îkâ ta-wî-mîtawâkât*
îkâ ta-wî-mîtawâkât-at
NEG FUT-PRSP-make.fun(VTA)-2SG>3SG.CNJ
'You are not to make fun of him' (PK)

Despite its oft-cited implications of desire, *wî-* is also able to occur on VIIs, whose inanimate (or even nonexistent) subjects lack any capacity for volition. For example:

190) a. *wâpahtam ôma â-wî-ihkinithik â-wî-itamothik itî mîskanaw îkotî*

wâpahtam ôma (k)â-wî-ihkin-ithik (k)â-wî-itamothi-k
see(VTI)-3SG.IND FOC CNJ-PRSP-exist(VII)-3`.CNJ CNJ-PRSP-connect(VII)-3`.CNJ

itî mîskanaw îkotî
place road(NI) there
'he saw this would be happening, where the railroad would connect' (PK)

b. *ôma kâ-wî-ispathik*

ôma kâ-wî-ispathi-k
DEM.PROX.INAN.SG CNJ-PRSP-happen(VII)-3SG.CNJ
'this week' (lit 'that which is going to happen') (SM) (Torkornoo et al. 2021)

The *wî-* prefix functions similarly in other the Western Cree languages, simultaneously indicating futurity and intentionality in varying measure depending on context. Wolvengrey (2006) notes that *wî-* in Plains Cree may be most accurately defined as marking Prospective aspect, that is, the notion that the event described is directly and pertinently related to a subsequent situation, while the other future tense prefixes (in Plains Cree, *ka-* and *ta-*) do not make any implication of a salient relation between events. For example:

191) a. *nika-nipân êkwa*

ni-ka-nipâ-n êkwa
1SG.IND-FUT-sleep(VAI)-1SG.IND and
'I will sleep now (but you guys can hang out if you want)' (Plains Cree)
(Wolvengrey 2006:406)

b. *niwî-nipân êkwa*

ni-wî-nipâ-n êkwa
1SG.IND-PRSP-sleep(VAI)-1SG.IND and
'I'm going to sleep now (so you guys will have to leave)' (Plains Cree)
(Wolvengrey 2006:406)

Tentatively, I have found no strong counter-evidence to this interpretation of *wî-* in the Woods Cree corpus; in this investigation, I will therefore be marking it as a Prospective aspect marker, rather than a tense marker. As such, the exact position in which *wî-* is affixed when multiple preverbs are present differs slightly from that of *na-*, *ka-*, *ta-*, and *kita-* (see Section 2.3.4.11). As a final note, I will also remark here that the semantic implications of volition and desire associated with *wî-* are generally stronger in Woods Cree than they are in Plains Cree, where *wî-* often functions almost indistinguishably from other tense markers. In this way, the Woods Cree *wî-* is perhaps closer in function to its equivalent in Swampy Cree, which Ellis (2015:167) also describes as denoting "intended or desired action".

2.3.4.4 Aspectual, Modal, and Manner/Direction Preverbs

In addition to tense preverbs, there exist a wide set of optional, non-lexical preverbs used to specify aspect, mode, manner, and direction. Aspectual prefixes occur after tense prefixes and before modal prefixes[53], while modal prefixes occur after aspectual prefixes and before lexical prefixes. Manner/Direction prefixes may occur immediately before or after modal prefixes; the relative order of the two is not fixed.

192) Aspectual Prefixes

ati-	*mâci-*	*pôni-*
'start'	'begin'	'stop

193) Modal Prefixes

nohtî-	*nihtâ-*	*kakwî-*
'need'	'be skilled in'	'try'

194) Manner/Direction Prefixes

pî-	*nitawi-*[54]	*isi-*
'come'	'go'	'thus'

Multiple prefixes from each of these classes may affix to a single verb stem:

195) Multiple Aspectual Preverbs

nikî-ati-mâci-ayânân
ni-kî-ati-mâci-ayâ-nân
1PL.IND-PST-start–start-be.there(VAI)-1PL.IND
'we started to live there' (PK)

196) Multiple Manner/Direction Preverbs

kâ-pî-isi-kiskisiyân
kâ-pî-isi-kiskisi-yân
CNJ-come-thus-remember(VAI)-1SG.CNJ
'the way I remember' (SM)

Aspectual, Modal, and Manner/Direction prefixes are strictly optional, however, there are many instances of prefix-stem collocations having become lexicalised, with their meanings shifting towards varying degrees of non-compositionality:

197) a. *nohtî-mîcisow*
nohtî-mîciso-w
be.hungry(VAI)-3SG.IND
's/he is hungry' (SM)
Note: Literally 's/he needs to eat'

b. *ati-itohtîw*
ati-itohtî-w
go.home(VAI)-3SG.IND
's/he is going home' (LR)
Note: Literally 's/he starts going'

[53] There are a small number of instances in the corpus of modal prefixes preceding aspectual prefixes (e.g. *ta-kakwî-kîsi-tipahamâ*(*hk*) - we are trying to finish paying our debt); however, the pattern of aspectuals preceding modals holds in the vast majority of cases

[54] *nitawi-* in particular is often extremely phonologically reduced, pronounced as *nta-*, *ntî-* or more rarely simply *tî-*

2.3.4.5 Reduplication

Partial reduplication may occur on any verb according to two possible patterns, so-called Weak (or Light) reduplication, which manifests as *C1ay-* (or *ay-* on vowel-initial stems), and Strong (or Heavy) reduplication, which manifests as *C1âh-* (or *âh-* on vowel-initial stems[55]). Both of these reduplication processes are used to mark aspect; Weak reduplication indicates progressive aspect, describing ongoing states or events, where Strong reduplication is used to indicate discontinuous aspect, describing events which occur, then cease to occur, and then occur again repeatedly (Ahenakew 1983):

198) a. *îskwâ î-ma-mîcisot mîsahamithiw ohkoma opahkîkinwaskisina.*
îskwâ î-ma-mîciso-t mîsah-amithiw
presently CNJ-RDPLW-eat(VAI)-3SG.CNJ mend(VTI)-3`.IND
o-hkom-a o-pahkîkinwaskisin-a
3SG.POSS-grandmother(NDA)-OBV 3SG.POSS-moccasin(NI)-PL
'While he eats his grandmother mends his moccasins.' (SM)

b. *kâ-pâh-pahkikawanithik pimiy*
kâ-pâh-pahkikawan-ithik pimiy
CNJ-RDPLS-drip(VII)-3`.CNJ fat(NI)
'the grease was dripping' (SL)

In addition to its aspectual meaning, Strong reduplication can also be used to describe spatial discontinuity, as in example 199. Strong reduplication can also be used distributively, to indicate that a state or event is occurring equally to all participants:

199) Spatial Discontinuity
a. *kâh-kiswâk îsa î-câh-cîstahikît*
kâh-kiswâ-k îsa î-câh-cîstahikî-t
RDPLS-be.close(VII)-3SG.CNJ reportedly CNJ-RDPLS-stitch(VAI)-3SG.CNJ
'the stitches were close to each other' (SM)

b. *kwayask mâh-mahkâhan*
kwayask mâh-mahkâhan-∅
properly RDPLS-wave(VII)-3SG.IND
'the waves were huge' (PK)

[55] More rarely, this can also surface as *ây-* (Ratt 2014:xvi)

200) Distributivity

a. *kâh-kisîtisahikîwak pâham*
kâh-kisîtisahikî-wak pâham
RDPLS-travel.quickly.by.vehicle(VAI)-3PL.IND perhaps
'They probably have fast dog teams' (PK)

b. *nîso asinîsa îtokî â-cîhkathât. mâh-misikitithiwa*
nîso asinîs-a îtokî kâ-cîhkath-ât
two stone(NA)-OBV DUB CNJ-put.in.fire(VTA)-3SG>3`.CNJ
mâh-misikiti-thiwa
RDPLS-be.big(VAI)-3`.IND
'He put two small rocks, I assume, into the fire. They were big rocks' (PK)

Weak reduplication may also be used in some instances for simple emphasis; however, this is less common and less systematic than its aspectual use:

201) a. *nikî-ati-ma-mîtho-ayânân*
ni-kî-ati-ma-mîtho-ayâ-nân
1PL.IND-PST-start-RDPLW-good-be(VAI)-1PL.IND
'we recovered fully'[56] (PK)

b. *î-kî-sa-sâkihât*
î-kî-sa-sâkih-ât
CNJ-PST-RDPLW-love(VTA)-3SG>3`.CNJ
'she was very fond of them' (PK)

The use of both reduplication types on a single stem is not attested in the corpus; however, when provided examples, my consultant preferred Weak reduplication to occur to the left of Strong reduplication (Solomon Ratt, personal communication via Arok Wolvengrey). This mirrors the pattern observed in previous research into reduplication in Plains Cree:

202) *nipa-pâh-pimohtân*
ni-pa-pâh-pimoht-ân
1SG.IND-RDPLW-RDPLS-walk(VAI)-1SG.IND
'I keep walking (all year round)' (Plains Cree) (Ahenakew 1983:371)

[56] This sentence is translated in the original text (ME-9) as "I recovered fully"; however, given its suffixal morphology and its discursive context, I have glossed it here as the exclusive first-person plural. In any case, this alternate analysis has no bearing on the reduplication

Generally, reduplication occurs after the modal preverb slot and before the lexical preverb slot. However, this is not always the case; Starks reports infrequent examples of reduplication attaching directly to the verb stem even when a lexical preverb is present, and there are several instances in the corpus of reduplication occurring before a modal preverb:

203) *onikahpisa nikî-âh-ati-onikânân*
onikahp-is-a ni-kî-âh-ati-onikâ-nân
portage(NI)-DIM-PL 1PL.IND-PST-RDPLS-start-portage(VAI)-1PL.IND
'We went over separate small places of portage' (SL)

2.3.4.6 Lexical Preverbs

Much like with nouns, there exists a wide range of lexical prefixes, often derived from nouns or verb Initials themselves, which serve to modify the semantics of the verbs to which they attach. They attach immediately prior to the verb stem, after any modal or inflectional prefixes. Just as lexical prenouns tend to act as adjective-like modifiers (see Section 2.3.1.6), lexical preverbs tend to serve an adverbial semantic function. As is the case with prenouns, lexical preverbs may be formed directly from nouns or verb initials by adding the 'particle suffix' *-i*; indeed, many preverbs are identical in both form and basic semantic character to their corresponding prenouns (see examples 247b-c and d-e):

204) a. *î-matwî-tîpwît*
î-**matwî**-tîpwî-t
CNJ-**loudly**-(VAI)-3SG.CNJ
's/he screams loudly, s/he is heard screaming' (PK, SM)

b. *î-ati-kaskitîwi-thikwaskwanithik*
î-ati-**kaskitîwi**-thikwaskwa-nithik
CNJ-start-**black**-be.cloudy(VII)-3`.CNJ
'there started to be dark clouds' (SM)

c. *kaskitîwi-kiyâskos*
kaskitîwi-kiyâskw-is
black-gull(NA)-DIM
'black tern' (SM)

d. *î-kî-misi-mâtot*
î-kî-**misi**-mâto-t
CNJ-PST-**big**-cry(VAI)-3SG.CNJ
'he cried so much' (SM)

e. *misi-kinîpik*
misi-kinîpikw
big-snake(NA)
'big snake' (SM)

2.3.4.7 Verb Classes

Verbs in Woods Cree may be divided into one of four classes based on their transitivity and the animacy of their arguments. Following the naming conventions of other Western Cree dialects,

these four subclasses are Inanimate Intransitive (VII), Animate Intransitive (VAI), Transitive Inanimate (VTI)[57], and Transitive Animate (VTA):

205) Inanimate Intransitive (VII)

a.	*nisiwanâtan*	b.	*kostâtikwan*
	nisiwanâtan-∅		kostâtikwan-∅
	be.spoiled(VII)-3SG.IND		be.frightening(VII)-3SG.IND
	'it spoils, it is destroyed' (LR)		'it is frightening' (LR)

206) Animate Intransitive (VAI)

a.	*nisiwanâtisiw*	b.	*kostâcîw*
	nisiwanâtisi-w		kostâcî-w
	be.spoiled(VAI)-3SG.IND		be.frightened(VAI)-3SG.IND
	's/he is spoiled, s/he perishes' (LR)		's/he is afraid' (LR)

207) Transitive Inanimate (VTI)

a.	*nisiwanâcihtâw*	b.	*kostam*
	nisiwanâcihtâ-w		kost-am
	spoil(VTI)-3SG.IND		be.afraid.of(VTI)-3SG.IND
	's/he spoils it, s/he destroys it' (LR)		's/he is afraid of it' (LR)

208) Transitive Animate (VTA)

a.	*nisiwanâcihîw*	b.	*kostîw*
	nisiwanâcih-îw		kost-îw
	destroy(VTA)-3SG>3`.IND		be.afraid.of(VTA)-3SG>3`.IND
	s/he destroys him/her' (PK)		's/he is afraid of him/her' (LR)

As demonstrated above, the most conspicuous formal identifier of a verb's class is its ending (and thus, its Final), with the Initial and Medial containing lexico-semantic information. The classes listed above each differ in the categories for which they inflect, as well as in the affixes with which those inflectional categories are indicated. However, all four follow the same Order system described in Section 2.3.4.2, and all four may be optionally affixed with the same sets of non-inflectional preverbs, following the same relative ordering.

[57] Wolvengrey (2011a) argues in favour of combining the categories of VTI and VAI into a category called V1, defined as containing all verbs with exactly one animate participant, in contrast to V0 verbs (or VIIs), which contain zero animate participants, and V2 (or VTAs), which contain two. I find considerable merit in this argument from a theoretical perspective, as VTI and VAI have extremely similar paradigms and syntactic behaviour, in contrast to VIIs and VTAs. However, I will continue to use the more orthodox tetradic subclass division throughout this section, as it is these designations which are by far more prevalent in existing Algonquianist literature.

2.3.4.7.1 Inanimate Intransitive Verbs (VII)

Inanimate Intransitive Verbs (or VIIs) are the least inflectionally complex subclass, encoding only number and obviation, and lacking imperative forms. Unlike in most other contexts, there is a morphological distinction in VIIs between singular obviative and plural obviative. VII stems may end in *â*, *î*, *i*, or *n*. Stems ending in *â*-, *î*-, and *i* follow the pattern illustrated in Table 18, while stems ending in *n* have their final *n* shift to an *h* in the Conjunct and Subjunctive (as may be seen in Table 19):

		Independent	Conjunct	Subjunctive
3SG		*mâthipathiw*	*î-mâthipathik*	*mâthipathiki*
3PL	Western	*mâthipathiwa*	*î-mâthipathiki*	*mâthipathikwâwi*
	Eastern		*î-mâthipathikwâ*	
3`SG		*mâthipathithiw*	*î-mâthipathithik*	*mâthipathithiki*
3`PL	Western	*mâthipathithiwa*	*î-mâthipathithiki*	*mâthipathithikwâwi*
	Eastern		*î-mâthipathithikwâ*	

Table 18. Paradigm for the vowel-final VII stem *mâthipathi-* (‘it goes badly’)

Some VII paradigms are even simpler than this; VIIs with no obvious referential subject (referred to in literature as Impersonal VIIs (Wolvengrey 2011a:110)) only inflect for obviation, having no number distinction. This is particularly common in VIIs relating to meteorological and geographical phenomena:

	Independent	Conjunct	Subjunctive
0	*kimiwan*	*î-kimiwahk*	*kimiwahki*
0`	*kimiwanithiw*	*î-kimiwanithik*	*kimiwanithiki*

Table 19. Paradigm for the *n*-final impersonal VII verb stem *kimiwan-* (‘it rains’)

VIIs may also be derived from VAIs using the Indefinite Actor suffixes *-niwan*, *-niwin*, or *-nâniwan*, with the resultant VIIs having the same paradigm as the above and the semantic character of an impersonal passive (e.g. Payne 2007:206). These suffixes are commonplace across the Western and Eastern Woods Cree, as well as in Swampy Cree. but less so in Plains Cree, where some Southern dialects have eschewed them almost entirely:

209)	a.	*nôtinitowak*	>	*nôtinitonâniwan*
		nôtinito-wak		[nôtinito-nâniwan]-∅
		fight.one.another(VAI)-3PL.IND		[fight.one.another-X](VII)-3SG.IND
		‘they fight one another’ (LR)		‘there is fighting going on’ (PK)
	b.	*ayamihâw*	>	*ayamihâniwin*
		ayamihâ-w		[ayamihâ-niwin]-∅
		pray(VAI)-3SG.IND		[pray-X](VII)-3SG.IND
		‘s/he prays’ (LR)		‘there is a (church) service’ (SM)

Aside from those noted in Table 18, there are several regional variations in VII paradigms between West and East. In addition to *-kwâ*, VII-stems ending with *â* or *n* have multiple possible proximate plural suffixes in the Eastern variety (*-kiht*, *-hki*), each of which is selected lexically, whereas the Western variety has only one (*-ki*) for all *â*-final stems. *î*-final VII stems also have this *-kiht* suffix, although it functions as the 3`PL Conjunct suffix:

210)	a.	*î-kâspâkiht*	b.	*î-kâspâki*
		î-kâspâ-kiht		î-kâspâ-ki
		CNJ-be.crisp(VII)-3PL.IND		CNJ-be.crisp(VII)-3PL.IND
		‘they (inan) are crisp’		‘they (inan) are crisp’
	c.	*î-ayahki*	d.	*î-ayâki*
		î-ayâ-hki		î-ayâ-ki
		CNJ-be.there(VII)-3PL.IND		CNJ-be.there(VII)-3PL.IND
		‘they (inan.) are there, they are present’ present’		‘they (inan.) are there, they are

2.3.4.7.2 Animate Intransitive Verbs (VAI)

VAI stems are slightly more complex, inflecting for the person, number, and obviation of their subject, as well as possessing an imperative paradigm.

	Independent	Conjunct	Subjunctive	Imm. Imperative	Del. Imperative
1SG	*ninipin*	*î-nipiyân*	*nipiyâni*		
2SG	*kinipin*	*î-nipiyan* *î-nipiyin*	*nipiyani* *nipiyini*	*nipi*	*nipihkan*
3SG	*nipiw*	*î-nipit*	*nipici*		
1PL	*ninipinân*	*î-nipiyâhk*	*nipiyâhki*		
12PL	*kinipi(nâ)naw*	*î-nipiyahk*	*nipiyahki*	*nipitân*	*nipihkâ* *nipihkahk*
2PL	*kinipinâwâw*	*î-nipiyîk*	*nipiyîko*	*nipik*	*nipihkîk*
3PL	*nipiwak*	*î-nipicik*	*nipitwâwi*		
3`	*nipithiwa*	*î-nipithit*	*nipithici*		

Table 20. Paradigm for the vowel-final VAI stem *nipi*-(‘to be dead’)

VAI stems may end in *o*, *i*, *î*, *â*, or *n*. Stems ending in *â* and *i* function as demonstrated in Table 20. *î*-final stems are slightly different; typically, the stem-final *î* changes to *â* in first and second-person Independent forms (although some *î*-final stems retain their final *î* in the speech of older speakers) (Starks 1992:272).

211) a. *nimâcân*
ni-mâcî-n
1SG.IND-hunt(VAI)-1SG.IND
‘I hunt’ (SL)

b. *nimâcîn*
ni-mâcî-n
1SG.IND-hunt(VAI)-1SG.IND
‘I hunt’ (SL)

o- and *ô*-final stems function according to the pattern outlined in Table 20; however, first and second-person suffixes in the Conjunct and Subjunctive (*-iyân*, *iyan*, *iyin*, *iyâhk*, *iyahk*, *iyîk*, etc.) are often produced with a *w* instead of an *y*, particularly in Eastern Woods Cree:

212) a. *î-mîcisoyân*
î-mîciso-yân
CNJ-eat(VAI)-1SG.CNJ
‘I eat’ (SM)

b. *î-mîcisowân*
î-mîciso-wân
CNJ-eat(VAI)-1SG.CNJ
‘I eat’ (SL, PK)

This systematic *w*~*y* variation can also be seen in first and second-person suffixes in the VTA paradigm (see, for instance, Table 23 in Section 2.3.4.7.4.2), although I have chosen to list only the *y* forms for each here.

n-final VAI stems differ substantially in their paradigm from other VAIs, as well as differing between Eastern and Western varieties (namely, in their first and second-person Conjunct and Subjunctive suffixes):

		Independent	Conjunct	Subjunctive	Imm. Imperative	Del. Imperative
1SG	West.	*nipimisinin*	*î-pimisiniyân*	*pimisiniyâni*		
	East.	*nipimisin*	*î-pimisinân*	*pimisinâni*		
2SG	West.	*kipimisinin*	*î-pimisiniyan*	*pimisiniyani*	*pimisini*	*pimisinihkan*
	East	*kipimisin*	*î-pimisinan*	*pimisinani*		
3SG		*pimisin*	*î-pimisih*(*k*)	*pimisihki*		
1PL	West.	*nipimisinân*	*î-pimisiniyâhk*	*pimisiniyâhki*		
	East.		*î-pimisinâhk*	*pimisinâhki*		
12PL	West.	*kipimisi*(*nâ*)*naw*	*î-pimisiniyahk*	*pimisiniyahki*	*pimisinitân*	*pimisinihkâ* *pimisinihkahk*
	East.		*î-pimisinahk*	*pimisinahki*		
2PL	West.	*kipimisinâwâw*	*î-pimisiniyîk*	*pimisiniyîko*	*pimisinik*	*pimisinihkîk*
	East.		*î-pimisinîk*	*pimisinîko*		
3PL		*pimisinwak*	*î-pimisihkwâw*	*pimisinihkwâwi* *pimisinitwâwi*		
3`		*pimisinithiwa*	*î-pimisinithit*	*pimisinithici*		

Table 21. Paradigm for the n-final VAI stem *pimisin-* (‘to lie down’)

Despite their name, VAI stems can often be semantically (if not syntactically) transitive, and, although they do not reflect them in their inflectional paradigms, can take optional objects:

213) a. *mâcîtan atihkwak*
mâcî-tan *atihkw-ak*
hunt(VAI)-12PL.IMP.IMM caribou(NA)-PL
‘Let’s hunt caribou’ (KN) (Torkornoo et al. 2021)

b. *nikî-nitawi-kimotinân iyakoni wîsakipicikwâsa*

ni-kî-nitawi-kimoti-nân	iyakoni	wîsakipicikwâs-a
1PL.IND-PST-go-steal(VAI)-1PL.IND	DEM.ANIM.OBV	crabapple(NA)-OBV

'we went to steal some of those crabapples' (SM)

As a final point, while most communities exclusively use *-cik* as the 3PL Conjunct suffix, certain Western communities use two different 3PL Conjunct suffixes, *-cik* and *-twâw*. This pattern of variation in the Conjunct 3PL suffix is also seen in other Western Cree dialects; in Plains Cree, Northern dialects tend to use *-twâw* while Southern dialects use *-cik*, and in Swampy Cree both *-cik* and *-twâw* are attested in the VAI paradigm (Ellis 2015:138). Whether there is any functional difference between *-cik* and *-twâw* in these communities is unclear; Howse (1844:201) notes *-twâw* and *-cik* as indicating "def. time" and "indef. time" respectively; however, in modern Woods Cree, Conjunct verbs in general are only optionally marked for tense. In any case, corpus data indicates that *-cik* is dominant in most communities, including those in the West, with only one modern speaker in the corpus (from an unidentified, but almost certainly Western[58], community) reliably using *-twâw* instead of *-cik*:

214) *môtha ayinânîwosâp î-itahtopiponîtwâw*

môtha	ayinânîwosâp	î-itahtopiponî-twâw
NEG	eighteen	CNJ-be.such.age(VAI)-3PL.CNJ

'[they are] under the age of eighteen years' (UL)

2.3.4.7.3 Inanimate Transitive Verbs (VTI)

VTI stems inflect for the same categories of subject person, number, and obviation as VAIs, and lack any object agreement. Similar to VAIs, although VTIs are referred to as transitive, and are always capable of accepting a syntactic object, it is not uncommon to see VTIs with either no overtly realised object or no object at all. Indeed, some VTI stems prefer this objectless function, and act semantically as if they were intransitive, despite inflecting according to the VTI pattern.

215) *kahkithaw awinak kî-mîthwîthihtam.*

kahkithaw	awinak	kî-mîthwîthiht-am
all	PRON.ANIM.SG	PST-be.happy(VTI)-3SG.IND

'Everybody was happy.' (PK)

Unlike VIIs and VAIs, VTI inflection differs little between East and West, aside from regular phonological processes:

[58] For example, this author exclusively used *ikwa* instead of *akwa*, did not reduce *kâ-* to *â-*, and did not use the *na-* future marker

	Independent	Conjunct	Subjunctive	Imm. Imperative	Del. Imperative
1SG	*nipîhtîn*	*î-pîhtamân*	*pîhtamâni*	-	-
2SG	*kipîhtîn*	*î-pîhtaman*	*pîhtamani*	*pîhta*	*pîhtamohkan*
3SG	*pîhtam*	*î-pîhta*(*hk*)	*pîhtahki*	-	-
1PL	*nipîhtînân*	*î-pîhtamâ*(*hk*)	*pîhtamâhki*	-	-
12PL	*kipîhtînaw*	*î-pîhtama*(*hk*)	*pîhtamahki*	*pîhtîtân*	*pîhtamohkâ* *pîhtamohkahk*
2PL	*kipîhtînâwâw*	*î-pîhtamîk*	*pîhtamîko*	*pîhtamok*	*pîhtamohkîk*
3PL	*pîhtamwak*	*î-pîhtahkwâ*(*w*)	*pîhtahkwâwi*	-	-
3`	*pîhtamithiwa*	*î-pîhtamithit*	*pîhtamithici*	-	-

Table 22. Paradigm for the VTI stem *pîht-* (‘to hear (s.t.)’)

VTI stems may also end in *â*, however, these *â*-final stems inflect identically to VAIs (see Table 20):

216) a. *nitâpacihtân*
ni-t-âpacihtâ-n
1SG.IND-LC-use(VTI)-1SG.IND
‘I use it’ (NC)

b. *kitâpacihtân*
ki-t-âpacihtâ-n
2SG.IND-LC-use(VTI)-2SG.IND
‘you use it’ (NC)

c. *âpacihtâw*
âpacihtâ-w
use(VTI)-3SG.IND
‘s/he uses it’ (NC)

2.3.4.7.4 Animate Transitive Verbs (VTA)

VTA stems have the most complex inflectional paradigms, with obligatory person, number, and obviation agreement for the subject and object. VTA agreement is subject to a person hierarchy, which follows the scheme 2>1>3>3`, where second-person is the highest rank on the hierarchy, and third-person obviative is the lowest (Wolvengrey 2011a:57). This person hierarchy can be used to divide all possible permutations of subject and object into three sets: the Local set, which involves only first and second-person arguments, the Mixed set, which involves first, second, and/or third-person argument, and the Third Person set, which involves only third-person arguments (Wolvengrey 2011a:65). Each of these sets may further be divided into two subsets; the Direct, in which the subject is higher on the person hierarchy than the object, and the Inverse, in which the object is higher. Inverse forms are typically given an overt theme marker (which differs between the Local, Mixed, and Third Person sets) to indicate that the participant which is lower on the person hierarchy is the subject, as otherwise, it is assumed by default that the participant which is higher on the person hierarchy is the subject. For example:

217) a. *kiwâpamin*
kiwâpamin
2SG>1SG.IND-see(VTA)-2SG>1SG.IND
'you see me' (NC)
Note: Expected hierarchical order 2>1 upheld

b. *kiwâpamitin*
ki-wâpam-**iti**n
1SG>2SG.IND-see(VTA)-1SG>2SG.IND
'I see you' (NC)
Note: Expected hierarchical order 2>1 not upheld, Inverse theme marker *iti-* added

218) a. *wâpamîw*
wâpam-îw
see(VTA)-3SG>3`.IND
s/he sees him/her/them (Obv.)' (NC)
Note: Expected hierarchical order 3>3` upheld

b. *wâpamikow*
wâpam-**iko**w
see(VTA)-3`>3SG.IND
's/he/they (Obv.) sees him/her' (NC)
Note: Expected hierarchical order 3>3` not upheld, Inverse theme marker *ikw-* added

It is impossible for both participants of a VTA to be on the same level of the person hierarchy; reflexives and reciprocals are both formed as derived VAI stems, not VTA stems, and if both participants of a VTA are third-person, one must be marked as obviative (see example 218):

219) a. *niwâpamison*
ni-wâpamiso-n
1SG.IND-see.self(VAI)-1SG.IND
'I see myself' (NC)

b. *wâpamihitowak*
wâpamihito-wak
see.one.another(VAI)-3PL.IND
'they see one another' (NC)

2.3.4.7.4.1 Compositionality

VTA inflectional affixes are not monolithic units; instead, they are each composed of multiple meaning-bearing components, combined in a mostly, but not entirely, predictable fashion. This is true of VII, VAI, and VTI affixes as well - compare, for instance, the 3SG and 3PL VTI forms provided in Table 22. However, this compositionality is most readily observable in the paradigms of VTAs, as their affixes fundamentally contain more information concerning person and number than those of the other verbal subclasses. A representative example of this may be found in the

plural object marking of Conjunct VTA suffixes. In the Mixed and Third Person sets, Conjunct suffixes which indicate a plural object are formed using the same suffix as would be used for the relevant person-number combination with a singular object with the addition of the suffix *-wâw* or *-ik*, with the latter of these affricating the /t/ in any *t*-final suffix to a /t͡s~t͡ʃ/. For instance:

220)	a.	*î-wâpamâyâ(hk)*	b.	*î-wâpamâyâhkwâw*
		î-wâpam-âyâhk		î-wâpam-[âyâhk-wâw]
		CNJ-see(VTA)-1PL>3.CNJ		CNJ-see(VTA)-[1PL>3.CNJ-3PL.OBJ]
		'we (excl.) see him/her' (NC)		'we (excl.) see them' (SM)
		Note: These affixes are described in Godden 2019		

221)	a.	*î-wâpamakiht*	b.	*î-wâpamakihcik*
		î-wâpam-akiht		î-wâpam-[akiht-ik]
		CNJ-see(VTA)-1PL>3.CNJ		CNJ-see(VTA)-[1PL>3.CNJ-3PL.OBJ]
		'we (excl.) see him/her' (PK, SL)		'we (excl.) see them' (NC)
		Note: These affixes are both described in Starks (1992:284-5)		

However, the distributions of these two possible plural markers among the VTA suffixes vary between communities, and even between individual speakers. For instance:

222)	a.	*î-wâpamâyîkwâw*	b.	*î-wâpamâyîkok*
		î-wâpam-[âyîkw-wâw]		î-wâpam-[âyîkw-ik]
		CNJ-see(VTA)-[2PL>3.CNJ-3PL.OBJ]		CNJ-see(VTA)-[2PL>3.CNJ-3PL.OBJ]
		'you (pl.) see them' (NC)		'you (pl.) see them' (NC)
		Note: 222a is found in Godden (2019), while 222b is found in Starks (1992:285)		

Indeed, in Eastern Woods Cree, some Conjunct plural object forms have neither suffix:

223)	a.	*î-wâpamak*	b.	*î-wâpamaki*
		î-wâpam-ak		î-wâpam-aki
		CNJ-see(VTA)-1SG>3SG.CNJ		CNJ-(VTA)-1SG>3PL.CNJ
		'I see him/her' (SL) (Starks 1992:284)		'I see them' (SL) (Starks 1992:285)

This phenomenon of Conjunct plural object marking is a representative example of the semi-irregular compositionality of VTA inflectional suffixes in general, and is far from the only such example. Indeed, in examining the full VTA paradigms provided in the following sections, I am confident that the informed reader will be able to identify other such patterns as well. However, in spite of these clear compositional elements, I will present VTA affixes in this section as holistic and indivisible units (as I have in previous sections). This is for two principal reasons. The first concerns the economisation of writing space, as the precise nature of the

compositionality of VTA affixes (and indeed, of inflectional affixes as a whole in Woods Cree) is a matter well warranting a thesis in its own right. The second of these concerns the computational model, as the representation of these suffixes as chunks more closely resembles how they are to be represented computationally in Section 3. In any case, a highly detailed account of the compositionality of VTA affixes in Plains Cree may be found in Wolvengrey (2011a).

2.3.4.7.4.2 Local Set

The first VTA affixation set is the Local Set, which uses the Inverse theme marker -*iti*:

	Direct		Inverse	
	Independent	Conjunct	Independent	Conjunct
2SG-1SG	*kiwâpamin*	*î-wâpamiyan*	*kiwâpamitin*	*î-wâpamitân*
2SG-1PL	*kiwâpaminân*	*î-wâpamiyâ*(*hk*)	*kiwâpamitinân*	*î-wâpamitâ(hk*)
2PL-1SG	*kiwâpaminâwâw*	*î-wâpamiyîk*	*kiwâpamitinâwâw*	*î-wâpamitakwâw*
2PL-1PL	*kiwâpaminân*	*î-wâpamiyâ*(*hk*)	*kiwâpamitinân*	*î-wâpamitâ*(*hk*)

Table 23. Paradigm of Independent and Conjunct forms for the Local set of the VTA stem *wâpam-* ('to see s.o.')

	Subjunctive		Imperative	
	Direct	Inverse	Immediate	Delayed
2SG-1SG	*wâpamiyani*	*wâpamitâni*	*wâpamin*	*wâpamihkân*
2SG-1PL	*wâpamiyâhki*	*wâpamitâhki*	*wâpaminân*	*wâpamihkâ*(*hk*)
2PL-1SG	*wâpamiyîko*	*wâpamita*(*h*)*kwâwi*	*wâpamik*	*wâpamihkîk*
2PL-1PL			*wâpaminân*	*wâpamihkâ*(*hk*)

Table 24. Paradigm of Subjunctive and Imperative forms for the Local set of the VTA stem *wâpam-* ('to see s.o.')

2.3.4.7.4.3 Mixed Set

The Mixed set is the largest VTA paradigm, and uses the theme marker -*ikw*:

		Direct		Inverse	
		Independent	Conjunct	Independent	Conjunct
1SG-3SG		*niwâpamâw*	*î-wâpamak*	*niwâpamik*	*î-wâpamit*
1SG-3PL	West	*niwâpamâwak*	*î-wâpamakwâw*	*niwâpamikwak*	*î-wâpamicik* *î-wâpamitwâw*
	East		*î-wâpamaki*		
1SG-3`		*niwâpamimâwa*	*î-wâpamimak*	*niwâpamikothiwa*	*î-wâpamimit*
2SG-3SG		*kiwâpamâw*	*î-wâpamat*	*kiwâpamik*	*î-wâpamisk*
2SG-3PL		*kiwâpamâwak*	*î-wâpamacik*	*kiwâpamikwak*	*î-wâpamiskwâw*
2SG-3`		*kiwâpamimâwa*	*î-wâpamimat*	*kiwâpamikothiwa*	*î-wâpamimisk*
1PL-3SG	West	*niwâpamânân*	*î-wâpamâyâ(hk)*	*niwâpamikonân*	*î-wâpamikoyâ(hk)*
	East		*î-wâpamakiht*		*î-wâpamiyamiht*
1PL-3PL	West	*niwâpamânânak*	*î-wâpamâyâhkwâw*	*niwâpamikonânak*	*î-wâpamikoyâhkwâw*
	East		*î-wâpamakihcik*		*î-wâpamiyamihcik*
1PL-3`		*niwâpamimânâna*	*î-wâpamimâyâ(hk)*	*niwâpamikonâna*	*î-wâpamikoyâhk*
12PL-3SG	West	*kiwâpamânaw*	*î-wâpamâya(hk)*	*kiwâpamikonaw* *kiwâpamikonânaw*	*î-wâpamikoya*(*hk*)
	East		*î-wâpama(hk)*		*î-wâpamitahk*
12PL-3PL	West	*kiwâpamânawak*	*î-wâpamâyahkwâw*	*kiwâpamikonawak* *kiwâpamikonânawak*	*î-wâpamikoyahkwâw*
	East		*î-wâpamahkwâw*		*î-wâpamitahkwâw*
12PL-3`		*kiwâpamimânawa*	*î-wâpamimâya(hk)*	*kiwâpamikonawa*	*î-wâpamikoyahk*
2PL-3SG	West	*kiwâpamâwâw*	*î-wâpamâyîk*	*kiwâpamik*(*o*)*wâw*	*î-wâpamikoyîk*
	East				*î-wâpamitîk*
2PL-3PL	West	*kiwâpamâwâwak*	*î-wâpamâyîkwâw*	*kiwâpamik*(*o*)*wâwak*	*î-wâpamikoyîkwâw*
	East		*î-wâpamâyîkok*		*î-wâpamitîkwâw*
2PL-3`		*kiwâpamimâwâwa*	*î-wâpamimâyîk*	*kiwâpamikowâwa*	*î-wâpamikoyîk*

Table 25. Paradigm of Independent and Conjunct forms for the Mixed set of the VTA stem *wâpam-* (‘to see s.o.’)

		Subjunctive		Imperative	
		Direct	Inverse	Immediate	Delayed
1SG-3SG		*wâpamaki*	*wâpamici*	-	-
1SG-3PL		*wâpamakwâwi*	*wâpamitwâwi*	-	-
1SG-3`		*wâpamimaki*	*wâpamimici*	-	-
2SG-3SG		*wâpamaci*	*wâpamiski*	*wâpam*	*wâpamâhkan*
2SG-3PL		*wâpamatwâwi*	*wâpamiskwâwi*	*wâpamik*	*wâpamâhkanik*
2SG-3`		*wâpamimaci*	*wâpamimiski*	*wâpamim*	-
1PL-3SG	West	*wâpamâyâhki*	*wâpamikoyâhki*	-	-
	East	*wâpamakihci*	*wâpamitahkwâwi*	-	-
1PL-3PL	West	*wâpamâyâhkwâwi*	*wâpamikoyâhkwâwi*	-	-
	East	*wâpamakihci*	*wâpamitahkwâwi*	-	-
1PL-3`		*wâpamimâyâhki*	*wâpamikoyâhki*	-	-
12PL-3SG	West	*wâpamâyahki*	*wâpamikoyahki*	*wâpamâtân*	*wâpamâhka(hk)*
	East	*wâpamahki*	*wâpamiyamihci*	-	-
12PL-3PL	West	*wâpamâyahkwâwi*	*wâpamikoyahkwâwi*	*wâpamâtânik*	*wâpamâhkahkik*
	East	*wâpamahkwayak*	*wâpamitahkwâwi*		*wâpamâhkanik*
12PL-3`		*wâpamimâyahki*	*wâpamikoyahki*	*wâpamâtân*	*wâpamâhkâhk*
2PL-3SG	West	*wâpamâyîko*	*wâpamikoyîko*	*wâpamihk*	*wâpamâhkîk*
	East		*wâpamitîko*	-	-
2PL-3PL	West	*wâpamayîkwâwi*	*wâpamikoyîkwâwi*	*wâpamihkwâw*	*wâpamâhkîkwâw*
	East	*wâpamayîko*	*wâpamitîko*	*wâpamik wâpamihkwâw*	*wâpamâhkik*
2PL-3`		*wâpamimâyîko*	*wâpamikoyîko*	-	-

Table 26. Paradigm of Subjunctive and Imperative forms for the Mixed set of the VTA stem *wâpam-* (‘to see s.o.’)

As Tables 25 and 26 demonstrate, much of the subdialectal variation in Woods Cree VTA affixes lies in the Mixed set, particularly with affixes involving a 1PL, 12PL, or 2PL actor. In these affixes, the Western form is typically an innovative one, closely resembling (or being identical to) the correspondent form in Plains Cree, whereas the Eastern form is typically more conservative of the original Proto-Algonquian (and thus often resembles the corresponding affix in modern Swampy Cree and Moose Cree). For example:

224) Proto-Algonquian *-akenči (1PL>3SG.CNJ) (Oxford 2023b)
Plains Cree: *-âyâhk*
Woods Cree:
Stanley Mission: *-âyâhk*
South Indian Lake:*-akiht*
Swampy Cree: *-akiht*
Moose Cree: *-akiht*

However, it should be noted that this geographic variation is not absolute. While the forms listed as 'Eastern' in Tables 25 and 26 are (at least in the corpus) only attested in the Eastern variety, the forms listed as 'Western' are not exclusive to the West, and are also attested in Eastern communities (albeit more occasionally):

225) *kâ-wâhkomâyâhkwâ*
kâ-wâhkom-âyâhkwâ(w)
CNJ-relate(VTA)-12PL>3PL.CNJ
'[those] who are related to us' (PK)
Note: *-âyâhkwâ*(*w*) is used for 12PL>3PL.CNJ, not the expected Eastern form *-ahkwâw*

2.3.4.7.4.4 Third Person Set

The Third Person set also uses the *ikw*- theme marker for Inverse forms, however, it differs from the Local and Mixed sets in other respects. Firstly, as mentioned, at least one of the arguments in the Third Person set must be obviative, so as to disambiguate subject from object. However, it is also possible for both arguments to be obviative. In these instances, one of the arguments (in the corpus, this is always the object) is considered a 'further obviative' (3``), triggering distinct inflection. These further obviative forms, particularly those with proximate subjects, are uncommon in Woods Cree (Starks 1987:339), although it is not clear if this is due to the form falling out of use, or simply a product of further obviatives being discursively uncommon in the first place; I leave this as a matter for future research. As might be predicted, when further obviative forms do occur, they occupy the lowest rung of the person hierarchy.

	Direct		Inverse	
	Independent	Conjunct	Independent	Conjunct
3SG-3`	*wâpamîw*	*î-wâpamât*	*wâpamikow*	*î-wâpamikot*
3PL-3`	*wâpamîwak*	*î-wâpamâcik*	*wâpamikowak*	*î-wâpamikocik*
3`-3``	*wâpamîthiwa*	*î-wâpamâthit*	*wâpamikothiwa*	*î-wâpamikothit*
3SG-3``	*wâpamimîw*	*î-wâpamimât*	-	-
3PL-3``	*wâpamimîwak*	*î-wâpamimâcik*	-	-

Table 27. Paradigm of Independent and Conjunct forms for the Third Person set of the VTA stem *wâpam-* (‘to see s.o.’)

	Subjunctive		Imperative	
	Direct	Inverse	Immediate	Delayed
3SG-3`	*wâpamâci*	*wâpamikoci*	-	-
3PL-3`	*wâpamâtwâwi*	*wâpamikotwâwi*	-	-
3`-3``	*wâpamâthici*	*wâpamikothici*	-	-
3SG-3``	-	-	-	-
3PL-3``	-	-	-	-

Table 28. Paradigm of Subjunctive and Imperative forms for the Third Person set of the VTA stem *wâpam-* (‘to see s.o.’)

2.3.4.7.4.5 Inanimate Actor

As shown throughout Section 2.3.4.7, although there is a verb class for intransitive verbs with (grammatically) inanimate subjects (VII), there is no subclass of transitive verb stems with inanimate subjects. The likely reason for this is pragmatic infrequency; events in which inanimate objects act on animate ones are relatively infrequent cross-linguistically, and tend to be seen as marked constructions (see, for example, Yamamoto 1999). It is perhaps for this reason that inanimate subject transitives are represented using a modified form of the VTA paradigm, known as the Inanimate Actor. Inanimate actor VTAs are indicated with the suffix *-iko*, which attaches directly to the verb stem, followed by the relevant person-number affix for the object (which is identical in form to the corresponding subject suffix from the VAI paradigm (as in Table 20).

	Independent	Conjunct	Subjunctive
INAN>1SG	*ninipahikon*	*î-nipahikoyân*	*nipahikoyâni*
INAN>2SG	*kinipahikon*	*î-nipahikoyan, î-nipahikoyin*	*nipahikoyani, nipahikoyini*
INAN>3SG	*nipahikow*	*î-nipahikot*	*nipahikoci*
INAN>1PL	*ninipahikonân*	*î-nipahikoyâhk*	*nipahikoyâhki*
INAN>12PL	*kinipahikonaw*	*î-nipahikoyahk*	*nipahikoyahki*
INAN>2PL	*kinipahikowâw*	*î-nipahikoyîk*	*nipahikoyîko*
INAN>3PL	*nipahikowak*	*î-nipahikocik*	*nipahikotwâwi*
INAN>3`	*nipahikothiwa*	*î-nipahikothit*	*nipahikothici*

Table 29. Inanimate Actor VTA paradigm for the stem *nipah-* ('to kill (s.o.)')

Structurally, Inanimate Actor forms are identical to typical VTA forms with 3` subjects; however, there is seldom any syntactic ambiguity between the two, as Inanimate Actor forms strongly prefer to have an overtly represented (inanimate) subject. This is, again, likely a product of the Inanimate Actor being a marked construction, with speakers feeling the discursive need to explicitly justify its use in any given circumstance. Although ostensibly a simple combination of person-number-animacy permutations, the Inanimate Actor paradigm has some pragmatic idiosyncrasies; for instance, no Subjunctive paradigm for the Inanimate Actor is attested in the corpus (although my consultant assures me that these forms (which I have reconstructed based on the equivalent paradigm in Plains Cree) are grammatical (Solomon Ratt, personal communication via Arok Wolvengrey)). The Inanimate Actor also has a notable preference for negative connotational semantics, and statistically favours 1SG and 2SG objects, while strongly disfavouring 3`objects.

226) a. *kâ-kî-isi-wanâhikoyân nikiskinwahamâkowin*
kâ-kî-isi-wanâh-ikoyân ni-kiskinwahamâkowin
CNJ-PST-thus-hinder(VTA)-INAN>1SG.CNJ 1SG.POSS-education(NI)
'my schooling has hindered me' (SM)

b. *kikiskisiwina î-kâsisiki ta-manisikoyan kitoskanihk*
kikiskisiwin-a î-kâsisi-ki ta-manis-ikoyan
memory(NI)-PL CNJ-sharp(VII)-3PL.CNJ FUT-slice(VTA)-INAN>2SG.CNJ

ki-t-oskan-ihk
2SG.POSS-LC-bone(NI)-LOC
‘your memories are sharp enough to slice through your bones’ (SM)
Note: Discussing the loss of a loved one

2.3.4.7.4.6 Unspecified Actor

Woods Cree has no overt passive-active voice distinction; however, when the focus of a VTA is on the object, with the subject being either unknown, irrelevant, or generic, a distinct subparadigm known as the Unspecified Actor is used. This Unspecified Actor is distinct from the Indefinite Actor discussed in Section 2.3.4.7.1; while the Indefinite Actor applies to VAIs and VTIs, changing the verb stem to a VII, Unspecified Actor verbs still function as typical VTA stems. An almost identical dichotomy between the Unspecified Actor and Indefinite Actor exists in Swampy Cree and Moose Cree (Ellis 2016); however, in Plains Cree, Unspecified Actor morphology may also be used with VAI and VTI stems[59]. The Woods Cree VTA Unspecified Actor may appear on Independent, Conjunct, or Subjunctive verbs; however, owing to its de-emphasis of the subject, Unspecified Actor forms in the Imperative do not exist.

	Independent	Conjunct	Subjunctive
X>1SG	*nipahikawin*	*î-nipahikawiyân*	*nipahikawiyâni*
X>2SG	*kipahikawin*	*î-nipahikawiyan* *î-nipahikawiyin*	*nipahikawiyani* *nipahikawiyini*
X>3SG	*nipahâw*	*î-nipahiht*	*nipahihci*
X>1PL	*ninipahikawinân*	*î-nipahikawiyâhk*	*nipahikawiyâhki*
X>12PL	*kinipahikawin(ân)aw*	*î-nipahikawiyahk*	*nipahikawiyahki*
X>2PL	*kinipahikawinâwâw*	*î-nipahikawiyîk*	*nipahikawiyîko*
X>3PL	*nipahâwak*	*î-nipahihcik*	*nipahihtwâwi*
X>3`	*nipahâthiwa*	*î-nipahimiht*	*nipahimihci*

Table 30. Unspecified Actor VTA paradigm for the stem *nipah-* (‘to kill (s.o.)’)

[59] Here, it serves largely the same semantic function as the Indefinite Actor, but does not convert the verb stem to a VII

227) a. *kâ-kî-sipwîhtahikawiyân iskonikanihk ohci*
kâ-kî-sipwîhtahikawiyân iskonikan-ihk ohci
CNJ-PST-take.away(VTA)-X>1SG.CNJ reserve-LOC from
'I was taken from my reserve' (SM)

b. *nipimahothikawinân*
ni-pimahoth-ikawinân
X>1PL.IND-haul(VTA)-X>1PL.IND
'We were being hauled along' (SM)

2.3.4.8 Marginal Paradigms

As discussed in Section 2.3.4.2, contemporary Woods Cree verbal morphology has simplified in many respects compared to its historical complexity. However, traces of three of these defunct paradigms still persist in the modern language, with varying degrees of conspicuity. These three are the Relational, a valency-altering paradigm used to include additional actors to VAIs and VTIs, the Dubitative, which functioned as the name implies, and the Preterit, a set of paradigms broadly used to indicate past-tense. I discuss each of these in detail in Appendix F; however, for the purposes of describing the language as it is used today, none of these paradigms are fully productive, with the Preterit and Relational being rare archaisms, and the Dubitative being entirely obsolete.

2.3.4.9 Diminituvisation

Diminutivisation for verbs occurs through largely the same process as it does for nouns; all /t/ segments are converted to /t͡s~t͡ʃ/, and the suffix *-si* is added to the verb stem. Semantically, diminutive verbs generally indicate that the state or action indicated by the verb is occurring to a lesser extent than would typically be expected, although the relationship between the meanings of some verbs and their diminutive counterparts is less compositional. When applied to VTI stems, diminutivisation converts the stem in question to a VAI.

228) a. *mâtow*
mâto-w
cry(VAI)-3SG.IND
's/he cries' (LR, SM, PK)

b. *mâcosiw*
mâcosi-w
cry.a.little(VAI)-3SG.IND
's/he cries a little' (LR, SM)

229) *â-kî-nipahcâsiyân*
(k)â-kî-nipahcâsi-yân
CNJ-PST-kill.a.little(VAI)-1SG.CNJ
'I killed a moderate amount of game' (PK)
Note: Compare to non-diminutive verb *nipahtâ-* ('to kill' (VTI))

Diminutive verbs are productive, but generally less frequent than diminutive nouns. Given their relative infrequency compared to diminutive nouns, their varying levels of semantic transparency, and the fact that they formally modify the stem, I will consider diminutive verbs to be distinct lexemes from their non-diminutive forms proceeding in this analysis.

2.3.4.10 Comitative

The Comitative is a derivational paradigm which converts VAI stems into VTAs using the circumfix *wîci-*⟩...⟨*-im*, modifying the verb's semantics to imply accompaniment or joint action. These forms are seen across both Western and Eastern varieties, and, much like the verbal diminutive, are used productively, if not exceptionally often (with 24 total instances in the corpus, attested on 10 different stems)

230) a. *wîci-kapîsîmin*
[wîci-kapîsi-im]-in
[COM-make.camp(VAI)-COM](VTA)-2SG>1SG.IMP.IMM
'Camp here with me' (PK)

b. *âskaw î-wîci-wacistonîmâcik cahcahkiwa*

âskaw	î-[wîci-wacistoni-im]-âcik	cahcahkiw-a
sometimes	CNJ-[COM-have.nest(VAI)-COM](VTA)-3PL>3`.CNJ	pelican(NA)-OBV

'they sometimes nest with pelicans' (SM)

Like the diminutive, I will consider comitative verbs in my analysis to be distinct lexemes from their non-comitative VAI stems, being the product of derivation.

2.3.4.11 Structural Schematisation of the Woods Cree Verb

Based on both corpus observations of prefix ordering, as well as on previous research on Plains Cree morpheme ordering by Wolvengrey (2012), Bakker (2006), and Harrigan et al. (2017), I have compiled the following schematisation for Woods Cree verbal morphology. In this schematisation, when present, the leftmost morpheme is always either the person-number prefix in the Independent (e.g. *ni-*, *ki-*) or the Conjunct prefix in the Conjunct (e.g. *î-*, (*k*)*â-*). After this comes the tense prefix (e.g. *ka-*, *ta-*, *kî-*), followed by the Aspectual prefix (e.g. *wî-*, *mâci-*, *pôni-*), Modal prefix (e.g. *nohtî-*, *nihtâ-*), and the Manner/Direction prefix (e.g. *pî-*, *nitawi-*), although there can be some variability in the relative ordering of these three elements, and multiple of each type may be added on a single verb stem. After these comes Reduplication, with Light reduplication preceding Heavy reduplication if both are present. After Reduplication is the lexical preverb, which immediately precedes the verb stem, and of which multiple may occur at once. Aside from the internal structure of the verb stem, suffixation in verbs is limited to person-number agreement, which is present for Independent, Conjunct, Subjunctive, and

Imperative verbs alike. Although, as mentioned, these inflectional suffixes are internally complex, for the purposes of this analysis, I will present them as holofixal units.

Person Marking (Independent) Conjunct Prefix (Conjunct)	Tense Prefix	Aspect Prefix (↺)	Modal Prefix (↺)	Manner/ Direction Prefix (↺)	Reduplication		Lexical Preverb (↺)	Stem			Person-Number Agreement Suffix
					Light	*Heavy*		*Initial*	*Medial*	*Final*	

Figure 7. Schematisation of the morpheme ordering in a theoretical maximally complex Woods Cree verb

With the exception of the stem and the person-number agreement suffix (which may be null), all of the cells in the above schematisation are potentially optional. Indeed, the corpus lacks any examples of a 'maximally' complex verb according to this template. In practice, verbs seldom have more than four prefixal elements at once; forms with five prefixes are attested in rare instances, and forms with six or more are not attested at all[60].

231) *â-kî-ati-isi-pî-kîwihocik*
(k)â-kî-ati-isi-pî-kîwiho-cik
CNJ-PST-start-thus-come-go.home(VAI)-3PL.CNJ
'they started coming back home' (PK)

Despite this lack of any attested 'maximally' complex verbs, the relative ordering of the elements in the above schema can be demonstrated using a set of attested non-maximally complex forms. A collection of these may be found in Appendix I.

[60] A corpus analysis of Plains Cree preverbs in Harrigan et al (2017:593) yielded identical results, with five preverbs of any kind being the maximum attested on any given stem (i.e. *ê-kî-wî-kakwê-miy-ôsîhtâcik* ('they had intended to try to make something well')).

3. COMPUTATIONAL MODELLING OF WOODS CREE

In this section, I will discuss the construction and internal workings of a finite-state-based morphological model of Woods Cree, adapted from an existing model of Plains Cree using the morphological information established in Section 2. I will also discuss the testing of this model using .yaml files and the text corpus as benchmarks, as well as the creation of two sets of regular expression rewrite rules to account for common morphophonological processes and orthographic variation. Finally, I will discuss the practical applications of this model through the creation of a morphologically intelligent online dictionary, *itwîwina*.

3.1 Finite State Morphological Models

A Finite State Machine (or FSM) is an abstract computational machine which may at any time be in one of a finite number of possible states, with this state being able to change according to inputs provided to the machine. A simple example of a finite state machine, as provided by Beesley & Karttunen (2003:1-3) is a typical light-switch, which may be flicked upwards to turn a light on, and downwards to turn the light off. The light may therefore be considered to have two possible states, on and off, with it being possible to provide the machine with input (that is, movement of the switch) to alter its current state. The status of the light connected to the light-switch is the output of the machine, with this output being determined by the input provided to it:

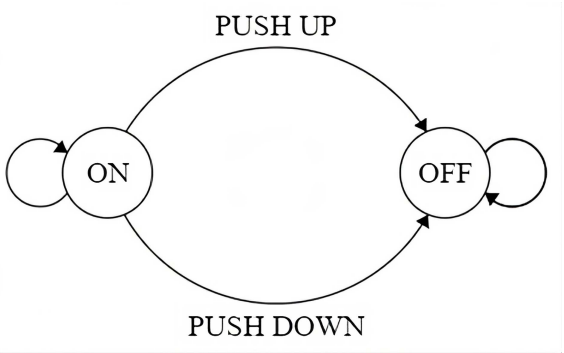

Figure 8. A Finite State model of a lightswitch (adapted from Beesley & Karttunen 2003:3)

A Finite State Transducer (or FST) is an FSM with two 'memory tapes'; that is to say, while a typical FSM can read input and provide appropriate output, an FST is also able to read output and reconstruct the necessary input. To return to the analogy of the light-switch, an FST would be able to determine if the light is on or off based on the position of the switch, but would also be able to determine the current position of the switch based on if the light is already on or off. This analogy describes an endlessly looping FSM, in which the state of the machine (and therefore its output) may be altered regardless of its current state. However, it is possible for an FSM to have terminal states which cannot be altered once they are reached. An analogy for this may be seen in an egg. The initial state of the egg is uncracked (or 'intact'), and it may persist in this state

indefinitely. However, once an input is given (striking the egg), the egg transitions to the terminal ‘cracked’ state, and cannot transition back to the intact state:

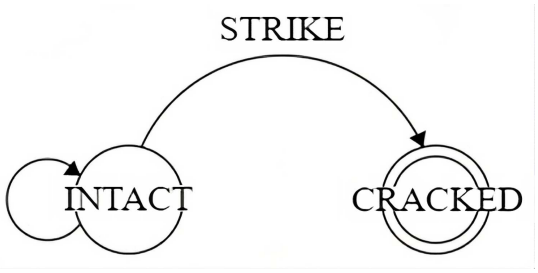

Figure 9. A Finite State model of an egg

FSTs have been used in natural language processing since the early 1980s (e.g. Koskenniemi 1983), chiefly for morphological modelling. A simplex stem is provided as an initial state for the machine, with multiple potential morphologically complex forms as terminal states. Inputs are given in the form of morphological categories (represented by pairs of underlying tags and surface representations). When an underlying tag is given as input, the state of the machine (the stem) transitions to a different state with a different surface realisation depending on the input. I provide a simplified example of an FST-based morphological model in Appendix J.

3.1.1 Existing Model(s)

FST-based morphological models already exist for various Algonquian languages, including Blackfoot (Kadlec 2023), Odawa (Bowers et al. 2017), Arapaho (Kazeminejad et al. 2017), and Plains Cree (Snoek et al. 2014; Harrigan et al. 2017). Of these, the Plains Cree model is the oldest and most elaborate, having been the subject of continual development since 2014 (Snoek et al. 2014). Based on this, and the extensive typological similarities between Plains Cree and Woods Cree, it is this model that will serve as the structural basis for the Woods Cree model. The Plains Cree morphological model was created following the Xerox-style definition of linguistic finite-state models (see Beesley & Karttunen 2003), implemented using FOMA (Hulden 2009) and HFST (Lindén et al. 2011), both C-based finite-state compilers designed to compile programs written in LEXC (a common FST formalism used for specifying morphosyntax across the lexicon in natural language processing) and XFScript. The morphotactic component of the model is written primarily using LEXC, with XFScript used to code rewrite rules. The LEXC component of the model is split internally into several submodels, one for each part of speech. Of these, the noun and verb submodels are further divided into stem files, which contain a full list of all relevant noun or verb stems, and affix files, which contain the relevant noun or verb affixes. After a lexeme from one of the stem files is provided morphology using its affix file, the resultant wordform is then parsed through the XFScript rewrite rules, which are common to all of the submodels, altering the wordform produced by the stem and affix files into the expected surface form. The general layout of this FST is governed by the file architecture of GiellaLT, a

platform of cross-language NLP infrastructure designed to accommodate FST-based morphological models for polysynthetic languages (Pirinen et al. 2023a; Pirinen et al. 2023b). The specifics of this file architecture are discussed in much greater detail in the official GiellaLT documentation[61]. However, within the Plains Cree model specifically, the internal structure(s) of each constituent file largely follow the LEXC and regular expression syntax demonstrated in the Appendix J.

One notable feature of the Plains Cree model is its liberal use of flag-diacritics, a type of tags used on both the surface form and analysis in LEXC to assign a given value to a lexeme only if certain preconditions are met (Beesley & Karttunen 2003:339-73). These flag-diacritics are used to limit the combinations of morphemes which the model considers legal, as well as to obligatorily co-ordinate discontinuous elements such as circumfixes. This is used in the Plains Cree model for phenomena such as noun possession. Much like in Woods Cree, plural possessors for nouns are indicated using a combination of prefixes and suffixes (see Section 2.3.1.3). However, possessive suffixes can only occur when an appropriate prefix is also present. For example, for the noun *cîman* (‘canoe’), the 1SG possessed form *nicîman* (ni-cîman) is allowed, as is the 1PL possessed form *nicîmaninân* (ni-cîman-inân), however, the form **cîmaninân* (cîman-inân) without the possessive prefix is ungrammatical. Flag diacritics can be used to prevent the model from producing the suffix *-inân* when the prefix *ni-* is not also present.

```
LEXICON Possessive_Prefixes
@P.person.NI@:@P.person.NI@ni%<  Noun_Stems ;

LEXICON Noun_Stems
cîman Possessive_Suffixes;

LEXICON Possessive_Suffixes
@U.person.NI@+Px1Sg:@U.person.NI@0
@U.person.NI@+Px1Pl:@U.person.NI@@C.sg@%>inân
```

Flag diacritics in LEXC are enclosed by the character `@`, and are, by default, automatically hidden from the surface representation of a lexeme when the model produces its final output. In the first lexicon, `Possessive_Prefixes`, the flag diacritic `P` is used, which sets a value for a variable assigned to the lexeme. In this instance, it creates a variable for the value of ‘person’, sets the value of this variable to `NI` (*ni-*), and also sets the surface form as `ni`. The flag diacritic itself is also encoded in the surface form, however, as mentioned, this will later be hidden in the final output. The model then proceeds to a noun stem (in this case, only `cîman` is available), and then goes to the `Possessive_Suffixes` lexicon. It then uses the flag diacritic `U`, which checks (and sets) values. If the value of ‘person’ is `NI`, it passes the check, and the value of

[61] https://giellalt.github.io/

+Px1Sg can be given, which has no surface representation (that is, no suffix). Alternatively, if the lexeme passes this check, the value of +Px1Pl can also be given, which has a surface representation of inân, added to the right of the noun stem. However, this check also means that the *-inân* suffix cannot be given unless the value of 'person' has been previously set to NI, which in turn means that it cannot be added unless the *ni-* prefix has already been attached to the noun stem in question. The flag diacritic U therefore acts as a gate for the attachment of the *-inân* suffix, allowing the grammatical forms *nicîman* ('my canoe') and *nicîmaninân* ('our (excl.) canoe'), but blocking the formation of the ungrammatical **cîmaninân*.

The specific details of the Plains Cree model's construction and internal structure are described in Snoek et al. (2014) and Harrigan et al. (2017). However, I will note here that the abbreviations used for certain morphosyntactic phenomena in the model differ somewhat from those which I have used in my interlinearised examples throughout Section 2, this discrepancy being carried forward from the Plains Cree model. As such, I have included a glossary of FST tag abbreviations on page xvi, which one may compare with the glossary of interlinear tags provided on page xii.

3.1.2 Descriptive and Normative Versions

The GiellaLT file infrastructure can be used to construct two distinct types of morphological analyser; a normative analyser and a descriptive analyser. A normative analyser only recognises and generates morphologically complex forms if they are exactly the same as the form expected by the model, abiding by all restrictions on affix-stem co-occurrences (such as those implemented using flag diacritics) and recognising only exact SRO spellings. A descriptive analyser recognises and generates morphologically complex forms even if they violate certain affix-stem co-occurrences or spelling conventions (see Section 3.3.8), as long as the requisite affixes and stems are present in the model. For instance, in the Plains Cree model, the noun stem file specifies that the noun *sîsîp* ('duck') requires the possessive theme suffix *-im* when it is possessed (e.g. *nisîsîpim*) (see Section 2.3.1.3.2), even though, in reality, the theme suffix may or may not occur in this context. However, because *sîsîp* is specified as a noun which takes *-im* in the noun stem file, the form *nisîsîp* (without the theme suffix) would be rejected by the normative analyser, as it violates one of the affix-stem co-occurrence restrictions specified by the flag diacritics. By contrast, the descriptive analyser is more lax; knowing that *ni* is an accepted prefix and *sîsîp* is an accepted stem, it would accept this form, even though it violates the aforementioned restriction. However, in order to indicate that this is not an expected form, the descriptive analyser would mark it with the tag +Err/Thm:

Input: sîsîp+N+I+Px1Sg+Sg
Normative: nisîsîpim
Descriptive: nisîsîpim, nisîsîp

Input: nisîsîpim

Normative: sîsîp+N+I+Px1Sg+Sg

Descriptive: sîsîp+N+I+Px1Sg+Sg

Input: nisîsîp

Normative: ???

Descriptive: sîsîp+N+I+Px1Sg+Sg+Err/Thm

These two analyser types are suited to different functions. The normative analyser is ideal for generating paradigms, as it only creates forms which would be strictly expected, as well as for spell-checkers. The descriptive analyser is ideal for parsing naturalistic texts, as it is able to allow for the irregularities in morphosyntax and orthography that are to be expected from natural language data, while still denoting unexpected forms with explicit tagging. Like the Plains Cree model, the Woods Cree model will also have functionality for both descriptive and normative versions.

3.2 .yaml files

Rather than manually verifying the grammaticality of all of its outputs, the outputs of the Plains Cree model are tested *en masse* using sets of testing files written in Yet Another Markup Language (or YAML). These testing files (henceforth .yaml files) are manually created sets of inflected wordforms, used as a gold standard for comparison with the forms and analyses produced by the FST. Each .yaml file contains the underlying tag-representation and surface form(s) of each expected inflected form for a given lexeme, as determined by a linguist. The lexemes selected as the bases for .yaml files are intended to be representative of distinct morphosyntactic or morphophonological classes, with at least one .yaml file of a representative lexeme for each such class being necessary to thoroughly assess the quality of the model. For example, a .yaml file for the English verb ‘cook’ might resemble the following

```
Lemma - cook

   cook+Inf:cook
   cook+3Sg:cooks
   cook+Past:cooked
   cook+Prp:cooking
```

In this .yaml file, the lemma `cook` has four expected forms in production, `cook+INF`, which surfaces as `cook`, `cook+3Sg`, which surfaces as `cooks`, `cook+Past`, which surfaces as `cooked`, and `cook+Prp`, which surfaces as `cooking`.

Using this .yaml file to verify the quality of a morphological model requires the use of a multifunctional script included in the GiellaLT infrastructure known as morph-test.py. This script first takes each underlying representation in the .yaml file as input for the word-form generator, and compares the output of the model with the corresponding word-form(s) provided in the .yaml file. Then, it takes each surface word-form in the .yaml file as input for the word-form analyser, and compares the output of the model with the corresponding underlying analyses provided in the yaml file. For instance, for the example above, the model would read the underlying tag representation `cook+INF` on the first line and provide a surface form (`cook`). Then, it would read the surface form also provided in the first line (`cook`) and provide an underlying analysis (`cook+INF`). Since both of these match the values given in the .yaml file, the morphological model may be considered to have 'passed' regarding that specific inflected form. The model would then analyse each other example in the .yaml file and run the same tests. To test the full breadth of the model, the model must therefore be tested on all parts of speech with distinct inflectional patterns.

For the Plains Cree model, the inflectional classes for which .yaml files were created were based on a combination of morphosyntax and phonological structure, the exact details of which are provided in Sections 3.3.3.1 and 3.3.4. However, broadly speaking, these subclasses are largely the same between Plains Cree and Woods Cree (see Wolvengrey 2011b). As such, the Plains Cree .yaml files were simply adapted for use with Woods Cree stems[62]. This process required roughly one-and-a-half weeks of manual work, compared to the potentially months-long process of creating a full set of .yaml files from scratch.

3.3 Adapting the Woods Cree Model

3.3.1 Adapting the Stem Lexicon

The first step necessary to adapt the Plains Cree model was to adapt the lexicon of possible stems, making it contain Woods Cree lexemic stems rather than Plains Cree ones. To do this, I used the 5858 entries of the *Colin Charles Cree Dictionary* to populate the stem lexicon. I used this dictionary rather than the *Westfall & Castel English-Cree Dictionary* for three reasons. Firstly, a digitised copy of the CCD already exists, whereas no such digitisation exists for the Westfall & Castel dictionary. Secondly, while the CCD typically provides only one entry for each lexeme, the Westfall & Castel dictionary has multiple entries for the inflected forms of most verbs, corresponding to whichever inflected forms were attested in the *Memoirs of the Elders*. Since the morphological model requires only stems for its lexicon, it would be necessary to

[62] Specifically, cognates of the same set of stems used for the Plains Cree .yaml files were used in the Woods Cree .yaml files. The structure of the paradigms represented in these .yaml files was also taken from the Plains Cree .yaml files, and were initially populated with potential inflected forms using the Plains Cree normative FST, modified with two broad Woods Cree phonological rules (ê > î, ý > th). These computationally generated .yaml files were then manually revised, as described in Sections 3.3.3 and 3.3.4)

simplify these hundreds of sets of inflected wordforms into discrete lexemic entries in order to use this dictionary in the model. Finally, as mentioned in Section 1.5, I[63] have been given explicit permission by the Lac La Ronge Indian Band, who own the rights to the CCD, to use this dictionary to develop digital language resources. As such, I have proceeded using only the CCD as a lexical basis for the model.

Several major alterations were necessary to adapt the CCD for use in the FST. Firstly, the spelling of each Cree entry needed to be converted from CMO to SRO, as the Plains Cree model was designed based on the latter. Given that both orthographies are standardised; this conversion was accomplished computationally using a small set of find-and-replace rules, subsequently verified by hand. The rules used were as follows: across all entry headwords, all instances of ⟨ch⟩ were replaced with ⟨c⟩, ⟨u⟩ was replaced with ⟨a⟩, ⟨a⟩ was replaced with ⟨â⟩, ⟨e⟩ was replaced with ⟨î⟩, and ⟨oo⟩ was replaced with ⟨ô⟩.

232) t**a**pisk**ooch** **e**-m**e**t**u**w**a**k**a**t**ach**ik
*t**â**pisk**ôc*** ***î**-m**î**t**a**w**â**k**â**t**âc**ik*
tâpiskôc î-mîtawâkât-âcik
like CNJ-play(VTA)-3PL>3`.CNJ
'it's like they are disrespecting them' (SM)

After this orthographic standardisation, each entry needed to be provided with a morphological analysis using the FST tags. This served primarily to allow any sets of multiple entries representing inflected forms of the same underlying stem to be automatically collapsed into a single entry, which was accomplished using a script running the Plains Cree FST. This also involved creating new entries for the 'Index' form of lemmas for which only inflected forms were originally given in the dictionary; for verbs, this Index form was the 3SG Independent, for independent nouns, the singular, and for dependent nouns, the singular Unspecified possessor form if possible, and the singular 1SG possessor form if not. The process of collapsing these entries together was accomplished with a script using the Plains Cree FST and a small set of rewrite rules simulating common Woods Cree phonological shifts (namely *ê* > *î* and *y* > *th*), followed by limited manual annotation. Additionally, if this script could match a Woods Cree entry in the CCD with an entry in the Plains Cree model (minus sound shifts), it automatically assigned this entry with the part-of-speech tag (see Sections 3.3.2 and 3.3.4), a stem, and analysis of its Plains Cree counterpart. Entries not recognised by the model were tagged manually; however, this amounted to only 1583 out of 6625 total entries. Ultimately, this roughly two-week process resulted in 217 additional entries, raising the total count to 6075.

In lieu of listing both grammatical and biological animacy for each noun entry in order to determine their eligibility for inflectional category suffixes (e.g. locative, distributive, obviative,

[63] Through my association with the Alberta Language Technology Lab

etc.), a distinct tagging system was used to indicate which nouns may take which category suffixes. This tagging system is described in Section 3.3.2. A less extensive tagging system was also used for verbs; plural-only verb stems (such as reciprocals) were tagged in the lexicon to indicate that only plural subject inflected forms should be generated for them[64]. Finally, this full lexicon was divided throughout the model into various, part of speech delineated 'stem' files; all noun stems were aggregated into a noun stem file, all verb stems into a verb stem file, and so on.

CMO	SRO	Analysis	Definition	Index Form	Class	Stem	Tags
achimew	âcimîw	`âcimîw+V+TA+Ind+3Sg+4Sg/PlO`	v.t. he/she speaks of him/her	âcimîw	VTA-1	âcim-	
mehchetiwuk	mîhcîtiwak	`mîhcîtiwak+V+AI+Ind+3Pl`	there are many of them (an.)	mîhcîtiwak	VAI-1	mîhcîti-	PluralOnly
noosemisk	nôsîmisk	`nôsîmisk+N+A+Sg`	a female beaver	nôsîmisk	NA-3	nôsîmiskw-	
usiniy	asiniy	`asiniy+N+A+Sg`	rock, stone (an.)	asiniy	NA-2	asiniy-	Locative
otehtehkosiwu	otîhtîhkosiwa	`mitîhtîhkosiw+N+A+D+Px3Sg+Obv`	his/her kidneys (an.)	mitîhtîhkosiw	NDA-2	-tîhtîhkosiw-	Locative; Unspecified
mistuhi	mistahi	`mistahi+Ipc`	a lot, much	mistahi	IPC		

Table 31. Sample entries from the CCD after having been adapted for use in the morphological model

3.3.2 Adapting Noun Affixation

Noun affixation differs little between Plains Cree and Woods Cree, and correspondingly, the noun affixation FST for the Woods Cree model is almost identical to its Plains Cree equivalent (described in detail in Snoek et al. 2014). Nouns in the model are divided into 18 subclasses (see Wolvengrey 2011b), with .yaml files created for each. These subclasses are determined by the phonological structure of the noun stem (and thus the morphophonological processes to which it is beholden) combined with its animacy and alienability. The phonological structure of a noun is indicated in its subclass label using a numerical identifier. Noun subclasses with the identifier 1 are typical consonant-final nouns, to which the morphemes contained in the model may be attached without any special morphophonological processes. Subclasses with the identifier 2 are semivowel final, and reduce stem-final *Vyi* or *Vwi* sequences caused by affixation to a lengthened singleton *V:*. Subclasses with the identifier 3 have an underlying final *Cw* sequence, in which the final /w/ only surfaces when an *a*- or *i*-initial suffix is attached; for the latter, the resultant /wi/ sequence is simplified to /o/. Subclasses with the identifier 4 are monosyllabic roots which must

[64] Although the descriptive analyser is still able to recognise singular subject forms of these stems, albeit with a `+Err` tag

occur with a number suffix (even in the typically-unmarked singular form) unless they are otherwise inflected in such a way as to make them polysyllabic. Subclasses with the identifier 4w are identical to this, aside from that the monosyllabic noun root has an underlying final /w/.

NA-1 (Consonant-final animate noun) - *kwâpahikan* (‘ladle’)

Regular animate noun, no special morphophonological processes

NA-2 (Semivowel-final animate noun) - *athapiy* (‘net’)

i-initial suffixes result in *Vyi* and *Vwi* sequences reducing to *V:*

- *athapiy-is* > *athapîs* (not **athapiyis*)
  net(NA)-DIM

NA-3 (*Cw*-final animate noun) - *amisk*(*w*) (‘beaver’)

i-initial suffixes are rounded to *o*, underlying *w* revealed on *a*-initial suffix

- *amiskw-is* > *amiskos* (not **amiskwis*)
  beaver(NA)-DIM
- *amiskw-ak* > *amiskwak* (not **amiskak*)
  beaver(NA)-PL

NA-4 (Vowel-final animate noun) - *aya* (‘living being’)

Stem must have a number-obviation suffix even when singular to avoid monosyllables

- *ay-a* > *aya* (but never **ay*)
  being(NA)-ANIM.SG

NA-4w (*wV*-final animate noun) - *ihkwa* (‘louse’)

As above, but the stem also contains an underlying final *w*

- *ihkw-a* > > *ihkwa* (but never **ihkw or *ihk*)
  louse(NA)-ANIM.SG
- *ni-t-ihkw-im* > *nitihkom* (not **nitihkwim*)
  1SG.POSS-LC-louse(NA)-POSS.THM

NAD-1 (Consonant-final dependent animate noun) - *mitâs* (‘trousers’)

NAD-2 (Semivowel-final dependent animate noun) - *mitihtihkosiy* (‘kidney’)

- *ni-tihtihkosiy-ihk* > *nitihtihkosîhk* (not **nitihtihkosiyihk*)
  1SG.POSS-kidney(NDA)-LOC

NAD-3 (*Cw*-final dependent animate noun) - *nîtim*(*w*) (‘opposite gendered cross cousin’)

- *n-îtimw-inân* > *nîtimonân* (not **nîtiminân*)
  1PL.POSS-cousin(NDA)-1PL.POSS

NAD-4w (*wV*-final dependent animate noun) - *nîskwa* (‘husband’s former wife’)

- *n-îskw-a* > *nîskwa* (not **nîskw*)
  1SG.POSS-EX.WIFE(NDA)-ANIM.SG
- *w-îskw-iwâw-a* > *wîskowâwa* (not **wîskwiwâwa*)
  3PL.POSS-EX.WIFE(NDA)-3PL.POSS-OBV

NI-1 (Consonant-final inanimate noun) - *cîmân* (‘canoe’)
NI-2 (Semivowel-final inanimate noun) - *maskihkiy* (‘medicine’)

- *maskihkiy-is* > *maskihkîs* (not **maskihkiyis*)
  medicine(NI)-DIM

NI-3 (*Cw*-final inanimate noun) - *kotawânâpisk*(*w*) (‘oven’)

- *kotawânâpiskw-ihk* > *kotawânâpiskohk* (not **kotawânâpiskwihk*)
  stove(NI)-LOC

NI-4 (Vowel-final inanimate noun) - *wâwi* (‘egg’)

- *wâw-i* > *wâwi* (but never **wâw*)
  egg(NI)-INAN.SG

NI-4w (*wV*-final inanimate noun) - *misko* (‘blood’)

- *miskw-i* > *misko* (but never **miskw*, **misk*, or **miskwi*)
  blood(NI)-INAN.SG

NID-1 (Consonant-final dependent inanimate noun) - *mitîh* (‘heart’)
NID-2 (Semivowel-final dependent inanimate noun) - *miskotâkay* (‘jacket’)

- *mi-skotâkay-ihk* > *miskotâkâhk* (not **miskotâkayihk*)
  X.POSS-jacket(NDI)-LOC

NID-3 (*Cw*-final dependent inanimate noun) - *miskîsik*(*w*) (‘eye’)

- *mi-skîsikw-is* > *miskîsikos* (not **miskîsikwis*)
  X.POSS-jacket(NDI)-DIM

NID-4 (Vowel-final dependent inanimate noun) - *nîki* (‘my home’)

- *n-îk-i* > *nîki* (but never **nîk*)
  1SG.POSS-home(NDI)-INAN.SG

For all inanimate nouns (NI or NDI)), number, possession, and locativity are encoded as affixal categories. Obviation is not encoded for NIs or NDIs, pending review of the productivity of inanimate obviative morphology. For all animate nouns (NA or NDA), number, possession, and obviation are encoded. An NA or NDA can also receive the locative suffix if it has the appropriate tag in the lexicon; however, if not, it cannot take the locative. The distributive, being a rare occurrence, is handled in the same way, with nouns known to be able to take the distributive suffix being tagged individually in the dictionary database. The vocative and absentative are not tagged at all; instead, attested vocative nouns are contained in a separate lexicon from other nouns, and absentative forms are listed as separate lexemes from non-absentative forms in the main stem lexicon (this convention being lifted directly from the Plains Cree model). Dependent nouns (NDI and NDA) are only stored in their possessed forms, but are otherwise identical in their affixation to non-dependent nouns of the same animacy value. Whether or not a given dependent noun can take the *mi-* prefix is also encoded through tagging in the dictionary database, instead of being left to be decided purely by the (grammatical) animacy of the noun in question. Throughout the model, the animacy value of each noun entry is

treated as an immutable and unchanging constant; nouns with variable animacy are listed as multiple separate entries.

All nouns may also be inflected for the diminutive, with diminutive forms existing for all possible number, possession, obviation, and locativity combinations. When inflected for the diminutive, nouns may take either the single diminutive suffix *-is* or the double diminutive *-isis*, and also receive a so-called 'diminutive trigger', `^DIM` (see Section 3.3.7). This trigger is used to facilitate Diminutive Consonant Symbolism in the post-hoc rewrite rules, after which point it is removed from the surface form by those rewrite rules.

Concerning lexical prenouns, only a small, systematically attested set from the Colin Charles Dictionary and the corpus are included, as well as a selection of forms from Plains Cree which have been adapted to Woods Cree phonology (taken from Wolvengrey (2011b)):

```
PN/askîwi+:askîwi PRENOUNS_BOUND "of the earth" ;
PN/atimo+:atimo PRENOUNS_BOUND "relating to dog, doglike" ;
PN/athakaski+:athakaski PRENOUNS_BOUND "wide" ;
```
etc.

Deverbal nominalisation is also not included in the model, nor are noun entries derived from verb roots linked in any way to the entries for the verbs from which they derive. In principle, modelling these processes using an FST is entirely possible (see, for instance, Kadlec 2023); however, I have excluded this functionality here due to time constraints.

3.3.3 Noun Affixation Path

To illustrate the basic structure of the noun model, I will provide here a truncated account of the path taken by a noun when it is analysed by it. To analyse a noun, the model begins in the noun affix file. It first determines if the noun in question is an independent noun, a dependent noun referring to a kin relation, or a dependent noun referring to a body part or other physical object; this latter distinction chiefly concerns the later assignment of the distributive and *mi-* possessor. The model determines the categories into which the noun in question falls based on directions assigned the noun entry in the stem lexicon. It then gives the noun in question a flag diacritic to indicate the category to which it belongs. After this, the model assigns a possessive prefix (the choices being moderated based on the categories detailed above); if the noun does not have a possessive prefix, the value for this slot is null. The noun is then assigned another flag diacritic to indicate its status of possession. After this, the model optionally assigns one or several lexical prenouns (which are enumerated in a separate lexicon file), and then takes the stem form of the noun in question from the noun stem file. Then, the model assigns the noun with a flag diacritic which indicates whether it habitually takes the possessive theme *-im* and whether it prefers double or single diminutive forms. After this, the model assigns the noun with a subclass tag

(+N+I, +N+A, +N+I+D, or +N+A+D) before adding the suffixes, with the noun being sent to a suffix lexicon corresponding to the subclass it was assigned in the previous step. The lexica containing the suffixes are 'gated', that is, whether or not a noun is sent through them is dependent on the flag diacritics assigned to that noun in the previous stages. First, the model may affix the possessive theme *-im* if the noun has a flag diacritic indicating that it takes the *-im* theme; if not, the noun skips this step. Then it assigns the diminutive suffix and trigger based on the previously assigned flag diacritic indicating which diminutive form is preferred; like the *-im* theme, this step may be skipped if the prerequisite flag diacritic(s) are not present. After this, the model assigns plural possessive suffixes. If the noun in question does not have a plural possessor, a null is assigned at this stage. Finally, the model assigns a number suffix, an obviative suffix, a locative suffix, or a distributive suffix; for singular nouns, this step assigns a null value.

Possessive Prefix	Lexical Prenoun (ᐎ)	Noun Stem	Possessive Theme, Diminutive Suffix	Plural Possessive Suffix	Locative, Distributive, Number, Obviation

Figure 10. A visual representation of the noun affixation path in the Woods Cree model

3.3.3.1 Noun .yaml Files

For the .yaml files for each noun subclass[65], I tested if the model could correctly produce the singular and plural form of the example noun in question, as well as all diminutive variants. For animate noun .yaml files, I also tested if the model could correctly produce the obviative and distributive forms, and for inanimates, I tested if it could produce the locative form. I also tested if the model could correctly produce each of these forms in tandem with possessive morphology (excluding singular and plural forms for third-person possessed animate nouns, as these must always be obviative). For Dependent nouns, the .yaml file contained only these possessed forms. Several examples of the structure of noun .yaml files may be found in Appendix K.

The noun stems used as representative lexemes for the .yaml files are identical to those provided in Section 3.3.2. In total, the .yaml files for each of the 5 non-dependent animate noun subclasses contained 24 possible form-analysis pairs, while the 5 non-dependent inanimate noun subclasses contained 38. The .yaml files for the animate dependent noun stems contained 23 pairs each (except for NDA-3, 4, and 4w, which have 20), while the inanimate dependent noun .yaml files contained 38 (except for NDA-4 (*-îskw*), which had 34). On testing the finished noun model with this set, the model attained a perfect score for each .yaml file, being able to both generate the expected output based on the provided analysis and generate the expected analysis based on the provided output for all form-analysis pairs throughout.

[65] https://github.com/giellalt/lang-cwd/tree/main/test/src/gt-norm-yamls

3.3.4 Adapting Verb Affixation

The Plains Cree verbal affixation model is able to analyse verbs in the Independent, Conjunct, Subjunctive (which it calls Future-Conditional (much like Wolvengrey (2011a)), and Imperative, as well as to model relevant prefixes for tense, aspect, modality, and manner/direction, as well as lexical preverbs and reduplication. It is then able to combine all of these permutations with all possible subject and object combinations and their relevant affixes. Given that all of these categories are identical at the basic, functional level in Woods Cree, I adopted the Plains verbal affixation system wholesale for use in the Woods Cree model. The only major paradigm that needed to be constructed from scratch was the VTA Inanimate Actor paradigm (see Section 2.3.4.7.4.5). The three marginal Woods Cree verb paradigms listed in Section 2.3.4.8, the Preterit, Relational, and Dubitative, were not modelled on the basis that they are no longer productive. The internal structuring of verb stems (concerning Initials, Medials, and Finals, as discussed in Section 2.3.4.1) was also not modelled, with lexemic stems acting as holistic and indivisible units. Concerning common verbal derivational processes, verbal diminutivisation was also not modelled, with attested verbal diminutives being listed instead as distinct lexemic entries. However, the comitative was modelled, although this was simply an opportunistic choice, as the Plains Cree model already contained code for comitative forms, operating identically to those in Woods Cree.

Concerning specific verbal affixes in the Plains Cree model which need to be changed for the Woods Cree model, aside from predictable phonological shifts (i.e. ⟨ê⟩ to ⟨î⟩ and ⟨y⟩ to ⟨th⟩), most differences were found in the VTA paradigms. For instance, Plains Cree VTA affixes with plural objects in the Mixed and Third Person sets typically end in *-ik*, whereas the corresponding affixes in Woods Cree typically end in *-wâw*; all such affixes thus needed to be altered for the Woods Cree model. The VTI paradigm for 3PL Conjunct verbs also differed in this respect, as did the Conjunct and Subjunctive VTAs involving a 2PL actor. The VTA, VII, and *n*-final VAI (or VAI-2) paradigms also demonstrated substantial subdialectal variation.

Subdialectal variation in affixes was represented using the tag `+Var/East`, which was added to the underlying form of any affix only found in the Eastern variety. For the purposes of the model, forms from the Western variety were considered the default, and were not explicitly tagged; this is due primarily to the lexical basis for the model, the *Colin Charles Cree Dictionary*, already representing the Western variety, as well as the fact that, in most affixes which have distinct Eastern and Western forms, it is the Western form which more closely resembles the equivalent in Plains Cree. For example, for the VAI *pimisin* (‘s/he lies down’), the expected 1SG Conjunct forms differ between the Eastern and Western varieties. In the West, *î-pimisiniyân* would be expected (see Table 21), almost identical to the form in Plains Cree (*ê-pimisiniyân*). However, in the East, *î-pimisinân* is the expected form, more closely resembling the equivalent in Swampy Cree and Moose Cree (*ê-pimisinân*), and by extension Proto-Algonquian (*pemihšinaːni). As

such, in the verb affix file, two different 1SG Conjunct suffixes were added for VAIs of this kind; *-iyân* being the default, and *-ân* being the Eastern form:

```
+1Sg:%>iyân
+1Sg+Var/East:%>ân
```

This regional variation was also reflected in the .yaml files, where both the Western and Eastern form were listed as expected wordforms:

```
PV/î+pimisin+V+AI+Cnj+1Sg: î-pimisiniyân
PV/î+pimisin+V+AI+Cnj+1Sg+Var/East: î-pimisinân
```

Ultimately, verbs were divided into 12 subclasses based on transitivity, animacy of referents, and stem-final consonant (which may determine various morphophonological processes related to suffixation). One or several .yaml files were created for each of these subclasses. The subclasses of VIIs are described in Section 2.3.4.7.1, and differ in the surface realisations of affixation accordingly. For VAIs, two subclasses exist; one for vowel-final stems (VAI-1)[66] and another for *n*-final stems (VAI-2). The differences in affixation between these two is described in Section 2.3.4.7.2. VTIs are also divided into two subclasses, consonant-final stems (VTI-1) and *â*-final stems (VTI-2), the differences between which are described in Section 2.3.4.7.3[67]. There is also a third subclass of VTI, VTI-3, defined as containing all *i*-final VTI stems, which inflect identically to *i*-final VAIs. In practice, this subclass contains only one stem, *mîci-* (‘to eat s.t.’).

Section 2.3.4.7.1 (vowel-final inanimate intransitive verb) - *miskwâ-* (‘to be red’)

- Typical VII stem, no special morphophonological processes

Section 2.3.4.7.1 (*n*-final inanimate intransitive verb) - *mâthâtan-* (‘to be bad’), *mîthwâsin-* (‘to be good’), *pimamon-* (‘to run along’)

- Stem-final /n/ is deleted in Conjunct forms with proximate subjects. In stems ending with the sequence /an/, the final /n/ is deleted and replaced with a /h/ in this context

î-pimamon-k	>	*î-pimamok* (not **î-pimamo**<u>**n**</u>**k)**
CNJ-run.along(VII)-3SG.CNJ		
î-mîthwâsin-k	>	*î-mîthwâsik* (not **î-mîthwâsi**<u>**n**</u>**k)**
CNJ-be.good(VII)-3SG.CNJ		
î-mâthâtan-k	>	*î-mâthâtahk* (not **î-mâthâta**<u>**n**</u>**k)**
CNJ-be.bad(VII)-3SG.CNJ		

[66] These are further divided in the LEXC code into VAIio (for stems ending in i or o), VAIae, VAIm (for a small number of m-final VAI stems), and VAIw_PL (for plural-only vowel-final stems)

[67] In the LEXC code, consonant-final VTI stems are referred to as VTIm, while *â*-final stems are VTIw. The small set of plural-only VTI stems are classed as VTIm_PL

VAI-1 (vowel-final animate intransitive verb) - *nîhithawî-* (‘to speak Cree’), *mâto-* (‘to cry’), *atoskî-* (‘to work’), *mîciso-* (‘to eat’)

- Typical VAI stem, no special morphophonological processes

VAI-2 (*n*-final animate intransitive verb) - *pimisin-* (‘to lie down’)

- Wholly distinct paradigm from VAI-1 (see Section 2.3.4.7.2)

VTI-1 (consonant-final transitive inanimate verb) - *nât-* (to get s.t.)

- Typical VTI paradigm, no special morphophonological processes

VTI-2 (*â*-final transitive inanimate verb) - *kîsihtâ-* (‘to complete s.t.)

- Distinct paradigm from VTI-1, identical to VAI-1

VTI-3 (*i*-final transitive inanimate verb) - *mîci-* (‘to eat s.t.’)

Irregular VTI stem, also identical to VAI-1

VTAs are divided into five subclasses. The first of these, VTA-1, contains most consonant-final VTA stems, to which affixes from the model may be attached without any special morphophonological processes. VTA-2 contains all VTA stems ending in the sequence *aw*; when *i*-initial suffixes whose *i* is a reflex of the Proto-Algonquian segment *e are attached to an *aw*-final stem, the resultant *awi* sequence reduces to a long *â.* However, when other *i*-initial suffixes are attached, the *Vwi* sequence is retained. VTA-3 contains all *Cw*-final VTA stems; all *i*-initial suffixes cause the sequence *Cwi* to reduce to *Co* in this class of stems[68]. VTA-4 contains all *t*-final VTA stems; in these stems, *i*-initial suffixes whose *i* is a reflex of the Proto-Algonquian *i or long *e· (which includes contemporary null suffixes such as the 2SG>3SG Immediate Imperative) cause the final /t/ to become an /s/. VTA-5 is similar to VTI-3 in that it contains only a single stem, the slightly irregular *it-* (‘to say to s.o.’); functionally, it is almost identical to VTA-4, albeit catered to the slight irregularities of the stem *it-*.

VTA-1 (most consonant-final transitive animate verbs) - *wîcih-* (‘to help s.o.’)

Typical VTA stem, no special morphophonological processes

VTA-2 (*aw*-final transitive animate verb) - *nitonaw-* (‘to look for s.o.’)

Some *i*-initial suffixes trigger reduction of the /awi/ to /a:/

- *nitonaw-ikow* > *niton**<u>**â**</u>kow* (not **niton**<u>***awi***</u>*kow*)
  look.for(VTA)-INAN>3SG.IND
  Note: *-ikow* descends from PrA *-ekwa (Oxford 2023b)

but

- *nitonaw-in* > *niton**<u>***awi***</u>*n* (not **niton**<u>***â***</u>*n*)
  look.for(VTA)-2SG>1SG.IMP.IMM
  Note: *-in* descends from PrA *-iro (Oxford 2023b)

[68] In the LEXC code, VTA-1, -2, and -3 are all classed simply as VTA

VTA-3 (*Cw*-final transitive animate verb) - *kîskisw-* (‘to play s.o. (i.e. in cards)’)

i-initial suffixes have their initial /i/ rounded to /o/

- *kîskisw-in* > *kîskis**o**n* (not **kîskis**wi**n*)
  play(VTA)-2SG>1SG.IMP.IMM

VTA-4 (*t*-final transitive animate verb) - *nakat-* (‘to abandon s.o.’)

Final /t/ is be spirantised to /s/ with the addition of certain suffixes

- *nakat-∅* > *naka**s*** (not **naka**t***)
  abandon(VTA)-2SG>3SG.IMP.IMM
  Note: -∅ descends from PrA *-i (Oxford 2023b)

but

- *ki-nakat-itin* > *kinaka**t**itin* (not **kinaka**s**itin*)
  1SG>2SG.IND-abandon(VTA)-1SG>2SG.IND
  Note: *-itin* descends from PrA *-eθeni (Oxford 2023b)

VTA-5 (*t*-final single syllable root transitive animate verb) - *it-* (‘to say to s.o.’)

Irregular VTA stem, final /t/ may be spirantised to /s/, as in VTA-4

3.3.4.1 Verb .yaml Files

Given the wildly differing extents of potential inflection among the paradigms of VIIs, VAIs, VTIs, and VTAs, each class had its own distinctly structured set of .yaml files, with each having .yaml files for up to several stems within that class in order to capture common morphophonological processes. Example sets of .yaml files for each of these subclasses (with accompanying structural descriptions) may be found in Appendix L. However, for all verb .yaml files, all anticipated inflected forms were listed multiple times in various tense, Order, and mood variants, including multiple common combinations of these variants. Since tense, Order, and mood do not differ between VII, VAI, VTI, and VTA, these tense-Order-mood combinations were identical between all .yaml files[69]. These preset combinations are provided below, giving an example using only the third-person singular forms of a VAI-1 for brevity:

Independent

Independent Present

`atoskîw+V+AI+Ind+3Sg:atoskîw`

Independent Past

`PV/kî+atoskîw+V+AI+Ind+3Sg:kî-atoskîw`

Independent Future Intentional[70]

`PV/wî+atoskîw+V+AI+Ind+3Sg:wî-atoskîw`

Independent Future Definite

`PV/ka+atoskîw+V+AI+Ind+3Sg:ka-atoskîw`

[69] The only exception to this is for VIIs, which do not have imperative forms

[70] As mentioned in Section 2.3.4.3, this *wî-* prefix is more accurately considered a prospective aspect marker, not a future marker. The labelling of it as a future intentional marker is an inheritance from the Plains Cree model.

Conjunct

Conjunct Present

`PV/î+atoskîw+V+AI+Cnj+3Sg:î-atoskît`

Conjunct Past

`PV/kî+atoskîw+V+AI+Cnj+3Sg:î-kî-atoskît`

Conjunct Future Intentional

`PV/wî+atoskîw+V+AI+Cnj+3Sg:î-wî-atoskît`

ka- Infinitive[71]

`PV/ka+atoskîw+V+AI+Cnj+3Sg:ka-atoskît`

ta- Infinitive

`PV/ta+atoskîw+V+AI+Cnj+3Sg:ta-atoskît`

Subjunctive

Future Conditional

`atoskîw+V+AI+Fut+Cond+3Sg:atoskîci`

Imperative

Immediate

`atoskîw+V+AI+Imp+Imm+2Sg:atoskî`

Delayed

`atoskîw+V+AI+Imp+Del+2Sg:atoskîhkan`

Conjunct forms in the .yaml files are all encoded using the *î-* prefix, rather than *kâ-*; however, this is purely for testing purposes, as *î-* and *kâ-* are both implemented in an identical fashion, making testing both redundant. Imperative forms are not coded to co-occur with tense or mood prefixes, as they are already subdivided into Immediate and Delayed tense variants. The .yaml files also lack tense prefixes for Subjunctive (or Future Conditional) forms, even though these forms are occasionally seen in the corpus.

Ultimately, the FST was able to attain a perfect score in generation and recognition for each of the .yaml files created for the four verb classes (see Appendix L).

3.3.4.2 Verbal Affixation Path

An overview of the verbal affixation path for the Plains Cree model may be found in Harrigan et al. 2017; this path is largely the same as that of the Woods Cree model. Nonetheless, I will provide a truncated account here. As with the noun model, the verbal affixation path begins in the verb affix file. Firstly, the model determines the verb's Order; that is, if it is Independent, Conjunct, Imperative, or Subjunctive/Future-Conditional, and directs to the relevant affixation lexicon accordingly. In doing so, the model also attaches a flag diacritic to the verb, indicating its Order.

[71] This category and the next are Conjunct verbs with a tense marker instead of a Conjunct prefix; referring to these forms as 'Infinitives' is again an inheritance of the Plains Cree model.

After the Order is determined, prefixes may be added. I will describe here the process for Independent verbs; however, the equivalent process for Conjunct verbs is identical in sequence, the difference lying only in the specific affixes used. To begin, the person prefixes *ni-* and *ki-* are added, both of which may be used for multiple possible person-number combinations. A null value may also be added at this stage for Independent verbs which lack a person-number prefix. All of these values are encoded with flag diacritics, as the presence or absence of a specific person-number prefix on a verb imposes restrictions on the suffixes with which that verb form may occur. After this, the verb is routed to a new lexicon which adds tense prefixes before proceeding to another lexicon called `VERBPREFIXES`. In the `VERBPREFIXES` lexicon, the verb has four options; it may either receive the comitative prefix *wîci-*, be routed to a new lexicon for reduplication, be routed to a new lexicon for other preverbs, or receive nothing. If the verb receives the comitative prefix, it is marked with a flag diacritic, sent to another lexicon to handle interactions between the *wîci-* prefix and other elements, and is then sent to the lexicon containing all verb stems. If the verb receives a reduplicative prefix, it is sent to a series of lexica which add one or several reduplicative elements before sending the verb to the main stem lexicon. If the verb receives any other prefixes, be they aspectual, modal, manner-directional, or lexical, it is sent to a continuation lexicon entitled `PREVERBS`. The model allows the aforementioned four prefix types to be attached in any order. This `PREVERBS` lexicon may apply elements recursively, allowing an indefinite number of prefixes to be added before moving to the stem lexicon.

After selecting a stem from the stem lexicon, the model adds tags to indicate the class of the verb in question (`+V+II`, `+V+AI`, `+V+TI`, or `+V+TA`), deciding this based on directions attached to each stem in the stem lexicon. Depending on the subclass to which the verb stem belongs, it is then sent to a series of continuation lexica containing suffixes. The first of these pertains to the *-im* suffix associated with the comitative; if the verb in question has the prefix *wîci-* (and therefore is marked with a comitative flag diacritic), it must take this *-im* suffix, and then be converted to a VTA. After this, the model again checks the Order of the verb form in question; in this case, since the verb form has a flag diacritic indicating that it is Independent, it is routed to a lexicon containing the Independent person-number suffixes associated with verbs of its class and it is given the tag `+Ind`. After this, the model uses checks for flag diacritics associated with person-number prefixes and routes the verb form to the appropriate lexicon of suffixes for the relevant Order which may co-occur with the person-number prefix already present. When these suffixes are added, they also add tags to indicate the relevant person-number combination which they represent. As mentioned, even for VTAs, these person-number suffixes are treated purely as chunks, and are the terminal element of any given verb form.

<table>
<tr><td>Person Marking (Independent)</td><td rowspan="2">Tense Prefix</td><td>Reduplication
----
Weak Strong</td><td rowspan="2">Preverb (↻)</td><td rowspan="2">Comitative Prefix</td><td rowspan="2">Stem</td><td rowspan="2">Comitative Suffix</td><td rowspan="2">Person-Number Suffix</td></tr>
<tr><td>Conjunct Prefix (Conjunct)</td></tr>
</table>

Figure 11. A visual representation of the verbal affixation path in the Woods Cree model

As mentioned, this process is essentially identical for Conjunct verbs, barring the fact that they are not sent to the person-number prefix lexicon, instead being sent to a lexicon for Conjunct prefixes, and that they are sent to a different set of suffix lexica. For the Subjunctive (or Future-Conditional) and Imperative, the verb form is not sent to any grammatical prefix lexica except for those for the comitative, for reduplication, and for (various) preverbs. In practice, this means that Subjunctive and Imperative verbs never receive person-number or tense prefixes in the model. Although, in the corpus, tense prefixes do occasionally occur with Subjunctive verbs, I have excluded these edge cases from the model to prevent over-generation.

3.3.5 Adapting Pronouns

Given their relatively small number and relative formal irregularity, instead of modelling pronouns using stems and affixes, I instead created a separate lexicon for pronouns which contained all expected forms paired with their full morphosyntactic analyses. As a consequence of this, each of these forms is deemed to be a separate lexical root; for instance, *ôma*, the proximate singular inanimate demonstrative, and *awa*, the proximate singular animate demonstrative, are considered two distinct lexical entries, connected only in the fact that they are both single, proximate, demonstrative pronouns. In addition to personal pronouns, demonstratives, and the various definite and indefinite pronouns discussed in Section 2.3.2, a number of common, non-inflecting interrogative pronouns were also included in the pronoun lexicon. Altogether, this lexicon contained 116 entries. Unlike the noun and verb lexica, which have multiple potential paths for affixation, the process of analysis for all pronouns simply involves the model checking if a given wordform is present in the pronoun lexicon, and if it is, providing the analysis listed in for that entry in the lexicon.

```
nîtha+Pron+Pers+1Sg:nîtha # ;
kîtha+Pron+Pers+2Sg:kîtha # ;
wîtha+Pron+Pers+3Sg:wîtha # ;
```
etc.

Since pronouns are analysed by the model simply by checking against a list of form-analysis pairs, no pronominal .yaml files were necessary (as there are no morphological processes to correctly or incorrectly apply).

3.3.6 Adapting Particles

Much like pronouns, all indeclinable particles, interjections, and particle phrases were contained in a separate lexicon. Within this lexicon, I defined three sublexica, these being `Particle` (for typical indeclinable particles) which assigned the tag `+Ipc`, `Particle/Interjection` (for interjections) which assigned the tag `+Ipc+Interj`, and `Particle/Phrase` (for space-separated particle phrases) which assigned the tag `+Ipc+Phr`. All particles were provided in a list, alongside directions for each to their proper sublexicon. Particle reduplication and diminutivisation were not modelled, as their productivity remains an open question; instead, particles attested with reduplicated or diminutive forms in the corpus and/or literature had these forms added to the particle lexicon as distinct entries. The only particles not included in this lexicon were numerals, which are contained in a separate lexicon so as to facilitate the eventual creation of a self-contained numeral model; however, this is a matter for future research.

Given that no morphological processes were modelled for particles, no .yaml files were constructed to test the quality of the model in recognising them.

3.3.7 Rewrite Rules

Most morphophonological phenomena are not coded directly in the affix files, and are instead applied separately using regular expression-based rewrite rules[72]. These rewrite rules are contained in a distinct XFScript file, and are applied only after all affixation has already taken place. As with so much else, these rules are adapted from the Plains Cree model.

To begin, the rewrite rules establish several general character sets, such as vowels, consonants, semivowels, and morpheme-boundary markers, which are to be used in later rules. These sets are:

```
define Bx [ %> | %< ] ;

define Ix [ i | i2 | ii2 ] ;

define Vowel [ a | â | i | î | o | ô ] ;

define Consonant [ c | h | k | m | n | p | s | t | w | y ];

define Gx [ w | y  ] ;
```

The character set `Ix` is worthy of further explanation. This set contains three characters; `i`, `i2`, and `ii2`. `i` represents the short high front vowel in general. `i2` represents specific instances of

[72] https://github.com/giellalt/lang-cwd/blob/main/src/fst/morphology/phonology.xfscript

this vowel which are descended from the Proto-Algonquian short vowel *e; this is relevant because these particular /i/ phonemes in modern Woods Cree trigger certain phonological processes not triggered by other /i/ phonemes (as mentioned in Section 3.3.4). This `i2` is present in the possessive theme marker, the locative suffix, the distributive, and plural possessive suffixes, as well as some person-number agreement suffixes for verbs. `ii2` is likewise used for long /i:/ segments in modern Woods Cree which derive from the PrA long vowel *e·, which behave similarly. This is not the only orthographic character with special marking for the rewrite rules; ⟨t⟩ is split into `t`, `t2` (which represents the optional linking consonant /t/) and `t3` (the /t/ at the end of VTA-4 stems), ⟨w⟩ is split into `w`, `w2` (a stem final /w/ which does not collapse when suffixes are added (e.g. *mo**w**-* ('to eat s.o.')) and `w3` (the /w/ which occurs as a third-person possessive prefix for certain dependent nouns), and ⟨n⟩ is split into `n`, `n2` (the /n/ at the end of first and second-person agreement suffixes for VAIs and VTIs) and `n3` (the final /n/ in certain VII and VAI-2 stems). These character tags do not represent distinct phonetic units; instead, they exist only to mark orthographic characters which are subject to specific morphophonological rules to which other instances of the same character representing the same phoneme in a different context are not. Similarly, the rewrite rules make use of so-called 'triggers' indicated with the character ^) to mark certain pieces of morphological information such as the diminutive. These, too, are simply multi-character indicators, not retained in the surface form.

After defining character sets, the first rewrite rule to be applied simplifies *Vw* and *Vy* sequences followed by short /i/ to *V:*, but only when the short ⟨i⟩ following the *Vw* or *Vy* sequence is an `i2`. After defining the three relevant cases where this rule may apply, the three resultant rules are concatenated into one single rule (`VGi2VVRule`) so as to be easier to compile later:

```
define VGi2aaRule [ [ a | â ] [ Gx ] %> i2 -> â %> ] ;
  -  mîskanaw%>i2hk > mîskan**âhk**  (not *mîskan**awihk**)
define VGi2iiRule [ [ i | î ] [ Gx ] %> i2 -> î %> ] ;
  -  nâpîw%>i2^DIMsis > nâp**îsis**  (not *nâp**îwisis**)
define VGi2ooRule [ [ o | ô ] [ Gx ] %> i2 -> ô %> ] ;
  -  sîwâpoy%>i2hk > sîwâp**ôhk**  (not *sîwâp**oyihk**)
define VGi2VVRule [ VGi2aaRule .o. VGi2iiRule .o.
                    VGi2ooRule ] ;
```

After this rule, there is a rule to simplify any sequences of two short ⟨i⟩s (separated by a morpheme boundary) into a single short `i`. This rule functions primarily to prevent NX-4s with `i`-initial suffixes from having illegal double-vowel sequences:

```
define iiRule [ i2 -> 0 || i Bx _ ] ;
  -  wâwi%>i2hk > wâw**ihk** (not *wâw**iihk**)
```

After this, there is a rule to simplify any sequence of `Cw` and `i` to `o` if a morpheme boundary intercedes them. This rule functions to cause *Cw*-final stems to round the vowel of *i*-initial suffixes, as is expected:

```
define Cwi2CoôRule [ w %> [ i | i2 ] -> %> o , w %> ii2 ->
%> ô  || Consonant _ ] ;
```

- `atimw%>i2^DIMs` > *acim**os*** (not **acim**wis***)

Similarly, the next rule simplifies any word-final *Cw* sequences to *C*:

```
define wFinalRule [ w -> 0 || Consonant _ .#. ] ;
```

- `atimw` > *atim* (not **atim**w***)

Next, a rule converts all `t` segments to `c` if the diminutive trigger `^DIM` is present, unless those `t` segments immediately precede `h` (and are therefore part of the digraph `th`):

```
define DimRule [ t -> c || _ \h ?* "^DIM" ] ;
```

- `thithîkisitân%>i2^DIMs` > `thithîkisicânis`
  (not **thithîkisitânis* or **chichîkisicânis*)

```
define t3DimRule [ t3 -> c || _ %> i2 "^DIM" ] ;
```

- `ot3%<othâkan%>i2^DIMs` > *ocothâkanis*
  (not **ocochâkanis* or **otothâkanis*)

After this, there is a rule to retain the linking consonant `t` in person prefixes only if the stem to which they attach is vowel-initial:

```
define t2epenthesisRule [ t2 -> 0 || _ %< Consonant ,, t2
-> t || _ %< Vowel ] ;
```

- `nit2%<api%>n` > *nitapin*
  but
- `nit2%<nipi%>n` > `ninipin` (not **nit**n**ipin*)

After this, there is a rule that allows an `h` to be epenthesised between any two vowels separated by an orthographic hyphen. The trigger `^EGLOT` is taken directly from the Plains Cree model; in practice, it represents an optional epenthetic `h`:

```
define h2epenthesisRule [ "^EGLOT" -> h || Vowel _ %< Vowel
] .o. [ "^EGLOT" -> 0 ] ;
```

- `î^EGLOT%<acimo%>t` > *î**h**acimot* (not **îacimot*)

The following rule causes `î` to become `â` when it immediately precedes a morpheme boundary and `n2`, allowing `î`-final VAI stems to have their final vowel shift to `â` in their first and second-person Independent forms:

```
define ee2aaVAIRule [ î -> â || _ %> n2 ] ;
```
- `nit2<nîhithawî>n2 >` *ninîhithaw**â**n* (not **ninîhithaw**î**n*)

A similar rule allows `a` to become `î` when it immediately precedes a morpheme boundary and `n2`, giving VTI stems the correct final vowel in first and second-person Independent forms:

```
define a2eeVTIRule [ a -> î || _ %> n2 ] ;
```
- `kit2<kiskîthihta>n2 >` *kikiskîthiht**î**n* (not **kikiskîthiht**a**n*)

The next rule converts the final `n3` in certain VII and VAI-2 stems to an `h` when it immediately precedes a `k`:

```
define n2hRule [ n3 -> h || _ %> k ] ;
```
- `î-<nîpin3>k >` *î-nîpi**h**k* (not **î-nîpi**n**k*)

Similarly, the following rule deletes final `n` segments entirely in the same context:

```
define n20VIIRule [ n -> 0 || _ %> k ] ;
```
- `î-<mispon>k >` *î-mispok* (not **î-mispo**n**k*)

The next rule converts the final `t3` in VTA-4 and VTA-5 to an `s` with it is followed by either an `i` or `ii2`, but retains it as a `t` when it is followed by an `i2`:

```
define t2sVTA4Rule [ t3 -> s || _ [ %> [ i | ii2 ] ] | .#.
,, t3 -> t || _ %> i2] ;
```
- `kit2<nakat3>in >` *kinaka**s**in* (not **kinaka**t**in*)
- `nakat3>i2kow >` *naka**t**ikow* (not **naka**s**ikow*)

The following rule converts the third-person dependent possessive prefix `w-` in on consonant-initial stems to `o-`, but retains it as `w-` elsewhere:

```
define w2oNDRule [ w3 %< -> 0 || .#. _ [ o | ô ] ,, w3 %<
-> o || .#. _ Consonant ] .o. [ w3 -> w ] ;
```
- `w3<spiton >` *ospiton* (not **w**w**spiton*)

but

- w3<îpit > *wîpit*

The next rule also deals with possessive prefixes for dependent nouns, lengthening the initial o on o-initial dependent nouns and deleting the short i in the first and second-person possessive prefixes (represented in the model as i4) when they co-occur:

```
define i2VPossNDRule  [ [ i4 %< o ] -> %< ô || .#. [ n | k
| m ] _ ,, i4 -> i || .#. [ n | k | m ] _ %< Consonant ]
.o. [ i4 -> 0 ] ;
```

- ni4<ohkom > *n**ô**hkom* (not **ni**o**hkom*)

The following rule lengthens any word-initial o if it is preceded by a possessive prefix:

```
define o2ooPossNounRule [ o -> ô || .#. [ [ n i ] | [ k i ]
| o ] [ t | t2 ] %< _ ] ;
```

- nit2<ospwâkan > *nit**ô**spwâkan* (not **nit**o**spwâkan*)

The final rule concerns reduplication; rather than listing all possible reduplicant prefixes as distinct affixes, I instead represent Weak reduplication in the model as d1a and Strong reduplication as d1âh and use a list of rewrite rules for each possible stem-initial consonant to cause d1 to assimilate to the relevant segment. If two reduplicative prefixes occur on the same stem, the initial consonant of the second is represented with d2, which functions identically:

```
define ReduplRule [ [ d1 | d2 ] -> c || _ [ \%< ]+ %< c ,,
[ d1 | d2 ] -> h || _ [ \%< ]+ %< h ,, [ d1 | d2 ] -> h ||
_ [ \%< ]+ %< h ,, [ d1 | d2 ] -> k || _ [ \%< ]+ %< k ,, [
d1 | d2 ] -> m || _ [ \%< ]+ %< m ,, [ d1 | d2 ] -> n || _
[ \%< ]+ %< n ,, [ d1 | d2 ] -> p || _ [ \%< ]+ %< p ,, [
d1 | d2 ] -> s || _ [ \%< ]+ %< s ,, [ d1 | d2 ] -> t || _
[ \%< ]+ %< t \h ,, [ d1 | d2 ] -> w || _ [ \%< ]+ %< w ,,
[ d1 | d2 ] -> y || _ [ \%< ]+ %< y ,, [ d1 | d2 ] -> {th}
|| _ [ \%< ]+ %< {th} ] .o. [ [ y2 | ý2 | y3 ] -> y || [ d1
| d2 ] ?* _ ] .o. [ [ d1 | d2 ] -> 0 ] .o. [ [ y2 | y3 ] ->
0 ] ;
```

- ***d1***ay2-<nipâw > ***na****-nipâw*
- ***d1***âh-<nipâw > ***nâh****-nipâw*
- ***d2***ay3-***d1***âh-<nipâw > ***na****-****nâh****-nipâw*

After these rewrite rules for morphological processes, I then define a three rules to alter any remaining special underlying characters into their default surface representation (e.g. `t2 > t`, etc.), to remove all remaining triggers, and to delete all boundary markers:

```
define deep2surfRule [ i2 -> i ,, ii2 -> î ,, n2 -> n ,, n3
-> n ,, t3 -> t ] ;
   - ni**t2**%<astotin > ni**t**%<astotin
   - ki<kâsîhkwâ>**n2** > ki<kâsîhkwâ>**n**
   - nîpi**n3** > nîpi**n**
   - naka**t3**>îw > naka**t**>îw
   - maskisin%>**i2**hk > maskisin%>**i**hk
   - nakas>**ii2**hkan > nakas>**î**hkan

define rmTriggerRule [ [ "^DIM" | "^COM" | "^IC" ] -> 0 ] ;
   - maskisin%>i**^DIM**s > maskisin%>is

define boundaryremover [ [ %< | %> ] -> 0 ] ;
   - maskisin**%>**is > maskisinis
```

Finally, I concatenate all of the above rules together in order of their application.

3.3.8 Spell-Relax

To allow the model to recognise common attested spelling variations and romanisations other than the SRO, a system of (optional) regular expression-based rules were adapted from the Plains Cree model. Collectively, these will be referred to as 'spell-relax' rules[73]. For instance, to model the orthographic free variation between ⟨ð⟩ and ⟨th⟩, I used the following rule:

```
{th} (->) ð ,,
   - wah**ð**aw > wah**th**aw
```

In prose, this rule reads that the character ⟨ð⟩ may be optionally substituted for character sequence ⟨th⟩ in any context for any wordform analysed by the model. This rule therefore allows the model to analyse any wordform written with ⟨ð⟩ (assuming that the wordform in question is otherwise analysable), even though the model exclusively uses ⟨th⟩ to represent this segment in the stem and affix files.

Firstly, a set of rules were written to account for the spelling conventions of the CMO, identical to those listed in Section 3.3.1. In regular expression syntax, they are written as follows:

[73] https://github.com/giellalt/lang-cwd/blob/main/src/fst/orthography/spellrelax.regex

```
{ch} (->) c ,,
   -  e-me**ch**isot > *î-mî**c**isot*
u (->) a ,
   -  **u**stotin > ***a**stotin*
a (->) â ,
   -  **a**poht**a**n > ***â**poht**â**n*
e (->) î ,
   -  **e**-m**e**chisot > ***î**-m**î**cisot*
{oo} (->) ô ,,
   -  m**oo**swu > *m**ô**swa*
```

The next set of spell-relax rules concerns vowel length diacritics; as mentioned in Section 2.1.1, both macrons and circumflexes are regularly used to represent long vowels in the SRO. Outside of the SRO, some authors also use acute accents or double vowels in order to mark length. However, authors often either omit vowel length markers altogether or mark length only in certain, contrastive cases. As such, the following rewrite rules allow the model to optionally recognise any graphemic short vowel (that is, any vowel with no diacritic) as a long vowel, permitting the model to analyse texts even if they are entirely devoid of vowel length marking. These rules also allow the model to optionally convert vowels with macrons, vowels with acute accents, and double vowels to vowels with circumflexes. I have also written rules for the inverse case, allowing the model to optionally convert vowels with circumflexes into bare vowels, allowing for some variation in long vowel placement. To provide an example of each for the grapheme ⟨a⟩:

```
â (->) a ,
   -  n**a**tam > *n**â**tam*
a (->) â ,
   -  nât**â**m > *n**â**tam*
ā (->) â ,
   -  n**ā**tam > *n**â**tam*
á (->) â ,
   -  n**á**tam > *n**â**tam*
{aa} (->) â ..
   -  n**aa**tam > *n**â**tam*
```

Spell-relax rules are also used to account for various deletion and reduction phenomena. For instance, the following rules allow the model to recognise a wordform even if any given word-medial short ⟨i⟩ segment is deleted or substituted for an apostrophe:

```
i (->) [..] || [ c | h | k | m | n | p | s | t | w | y ] _
[ c | h | k | m | n | p | s | t | w | y ] ,,
i (->) [ %' | ‘ | ’ ] ,
  - tânsi > tân**i**si
  - tân’si > tân**i**si
```

In theory, similar rules could also be written to account for the deletion of short /a/ and /o/; however, including optional deletion and substitution rules for all three of these vowels causes the model to become overly lenient in its recognition, generating many more analyses for any given wordform than is practical. As such, I limited these vowel deletion rules to short /i/, which in any case is the vowel that is most frequently deleted. However, a similar rule was written for the insertion (or deletion) of /h/ before any stop consonant, as pre-aspiration of this kind is subject to a great deal of regional and idiolectal variation:

```
h (->) 0 || [ a | i | o | â | î | ô ] _ [ c | k | p | t ]
,,
  - sikiw > si**h**kiw
  - makîsîs > ma**h**kîsîs
[..] (->) h || [ a | i | o | â | î | ô ] _ [ c | k | p | t
] ,,
  - sîsî**h**p > sîsîp
```

A set of rules was also included to account for the non-standard orthographic representation of allophonic voicing. For instance:

```
k (->) g,
  - sâ**g**ahikan > sâ**k**ahikan
p (->) b,
  - sî**b**iy > sî**p**iy
t (->) d,
  - i**d**a > i**t**a
c (->) j,
  - kih**j**i > kih**c**i
```

Various rules were also written to account for phonological reductions, including cluster reduction and various vowel-semivowel reductions and substitutions. For instance:

```
[..] (->) h || _ .#. ,,
  - nipâ**h** > nipâ
```

```
{hk} (->) 0 , {sk} (->) s || _ .#. ,,
  - maskisini > maskisini**hk**
  - piyis > piyis**k**
[ {aw} | {iw} ] (->) {ow},
  - nîhith**ow** > nîhith**aw**
  - omasinahikan**ow** > omasinahikan**iw**
{ow} (->) {iw},
  - mîcis**iw** > mîcis**ow**
{ow} (->) {aw},
  - isith**aw**îw > isith**ow**îw
[ {iyi} | {iwi} ] (->) [ î | i ],
  - **î**sâhok > **iyi**sâhok
{nwa} (->) {no} ,,
  - kiski**no**hamâkîw > kiski**nwa**hamâkîw
{iy} (->) î || _ .#. ,,
  - maskos**î** > maskos**iy**
```

A set of rules were also included to allow for variance in hyphenation, allowing the model to recognise wordforms without hyphens, with spaces instead of hyphens, or with excessive hyphens:

```
[ %- ] (->) [ %- ]+,
  - î**--**mîcisot > î**-**mîcisot
[ "@U.joiner.NULL@" %- ] (->) 0,
  - îmîcisot > î**-**mîcisot
%- (->) [ %  | %  | %  ]+,
  - î**_**mîcisot > î**-**mîcisot
```

Finally, various rules were included to account for potential English orthographic influences, such as the use of ⟨ee⟩ for ⟨î⟩ or the use of ⟨tch⟩ or ⟨ts⟩ for ⟨c⟩:

```
[ i | î ] (->) {ee},
  - **ee**-m**ee**tchisot > **î**-m**î**cisot
c (->) {tch},
  - ee-mee**tch**isot > î-mî**c**isot
c (->) {ts},,
  - ee-mee**ts**isot > î-mî**c**isot
```

Although the above rules are able to account for the majority of common orthographic variations in Woods Cree, I have intentionally excluded several known phonological phenomena to avoid

over-generation or over-leniency in analysis on the part of the model. These phenomena are primarily common phoneme substitutions, such as the occasional free variation between /n/ and /m/ and /w/ and /j/, as well as intervocalic consonant deletions (see Section 2.2.3).

3.4 Evaluating the Model

3.4.1 Analysing the Corpus

Having already tested the morphological model on curated sets of paradigms in the form of the .yaml files, the next logical step was to test its ability to analyse naturalistic running text. For this purpose, the morphological model was applied to the contents of the text corpus. This served two purposes; firstly, as mentioned, it would demonstrate the model's capacity for analysing the contents of a text without being specifically adapted for it (as was done to perfect the .yaml files), and secondly, it would provide morphosyntactic analyses for (at least a portion of) the text corpus, as the results of the model for each lexeme could be added directly to the .vrt files containing the corpus' contents. To simulate a truly 'blind' test, the contents of the text corpus were not modified orthographically prior to being run through the model; all variation in spelling conventions between texts was retained. This served the dual purpose of testing the effectiveness of the spell-relax rules (as described in Section 3.3.8).

To analyse the corpus, every wordform was first analysed by the normative model, with the descriptive model then analysing any remaining unanalysed forms. Then, a Plains Cree-based constraint grammar parser (Schmirler et al. 2018) was used to facilitate a limited degree of syntactic disambiguation. Overall, this method was able to analyse 40,250 out of 63,968 Cree word tokens, or 62.9%, and 8,667 out of 18,647 Cree word types, or 46.5%. The majority of words which could not be analysed were simply stems which were not represented in the FST lexicon; this much is demonstrated by the discrepancy between the proportion of tokens recognised by the model compared with types.

3.5 Practical Application

Having completed the morphological analyser(s), and having access to a sizable digitised lexicon of Woods Cree words with English translations, one immediately obvious practical application of this model was to create a morphologically intelligent online dictionary which uses the FST to recognise and generate inflected wordforms. As with so many other aspects of this thesis, a similar tool already exists for Plains Cree, named *itwêwina* (lit. 'words') (https://itwewina.altlab.app/) (Arppe et al. (2018; 2022; 2023). The underlying computational machinery for *itwêwina* is entirely language-neutral[74], and intelligent dictionaries using the same framework have been created for unrelated languages such as Tsúùtʾínà (https://gunaha.altlab.dev/). The only prerequisites for its use are a morphological model and a

[74] See the official documentation here: https://morphodict.readthedocs.io/en/latest/

lexicon of stems. Having both of these things for Woods Cree, this framework was adapted to create an online dictionary of Woods Cree, linked below:

https://itwiwina.altlab.dev/

Search queries on *itwîwina* can be entered in English or in Cree, with no need for the user to specify. The search functionality of *itwîwina* also makes use of word2vec, a semantic model for English which relies on the multi-billion word Google News Corpus (Mikolov et al. 2013). This model is able to represent the distribution of any given word in the corpus using a set of numerical values known as vectors, and is able to determine the (semantic) similarity of any two words in this corpus by comparing these vectors using cosine distance. The precise internal mechanics of this system are described in more detail in Jurafsky & Martin (2016), however, the practical output is that word2vec can rank every English word in any given text based on how closely related they are to an input word (provided that both words occur at least once in the Google News Corpus). As such, for *itwîwina*, when a user enters an English word as a search query, we are able to use word2vec to compare the semantic relatedness of the search query word to every (English) word in the definitions of every entry in the database and subsequently return the top ranking words as search results, in addition to any exact word matches. Thus, if a user searches the English word ‘blizzard’, while the first search result is an exact match for the word (e.g. *kisinâw* (‘it is cold (weather) blizzard’)), the following results are all semantically related entries not containing the word (e.g. *kôna* (‘snow’), *kaskatin* (‘it freezes’), etc.).

This semantic search functionality has been lifted entirely from the Plains Cree *itwêwina*; the implementation and functionality of the system in *itwêwina* is described in detail in Arppe et al. (2023). Since word2vec relies on data from English corpora, related entries can only be returned for search queries entered in English.

The search functionality of *itwîwina* is also able to recognise inflected wordforms (including those spelled in a nonstandard fashion) using the generative morphological model, and is able to link these inflected forms back to the appropriate entry. For instance, if a user entered *nitastotin* (nit-astotin, ‘my hat’), they would receive the entry *astotin* (‘hat’) as the top search result:

Figure 12. Search results on *itwîwina* for the query ‘nitastotin’

Alternatively, if they entered *întawîthitakwâw* (a nonstandard spelling of *î-nitawîthihtahkwâw* (î-nitawîthiht-ahkwâw, ‘they want it’)), they would be receive the entry *nitawîthihtam* (‘s/he wants it’) as the top search result:

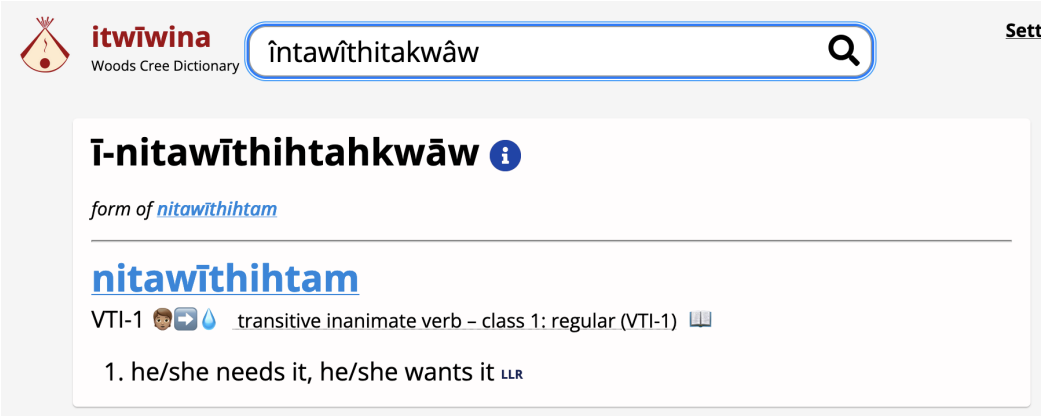

Figure 13. Search results on *itwîwina* for the query ‘întawîthitakwâw’

Using the normative model, *itwîwina* is also able to generate paradigms for each entry in the dictionary automatically, presenting these paradigms when a user clicks on an entry. The forms generated in these paradigms are identical to those tested using the .yaml file(s) for the part-of-speech of the relevant entry. Thus, if a user clicks on the entry for *kimiwan* (‘it rains’), they are shown the following:

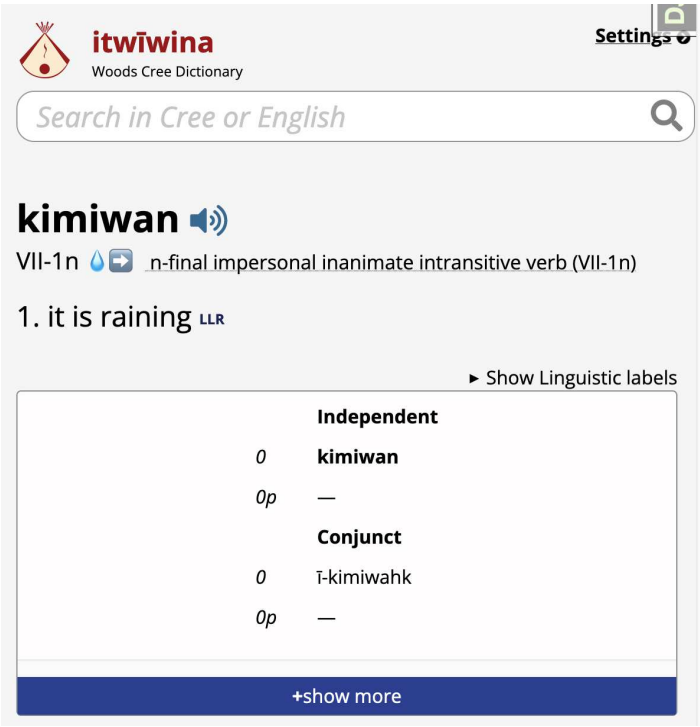

Figure 14. Truncated paradigm for the verb *kimiwan* on *itwîwina*

4. CONCLUDING THOUGHTS

In this section, I will discuss various possibilities for future research concerning Woods Cree documentation, as well as the manifold potential applications of the morphological model discussed in Section 3. Finally, I will provide a summary and conclusion to the research endeavours detailed thus far.

4.1 Future Research

As I have indicated in Sections 1 through 3, the research presented here is merely an introductory step in the improvement of the documentation of the Woods Cree language. For instance, although I have provided general overview of subdialectal differences between Eastern and Western Woods Cree from the corpus across Sections 1 and 2, I have yet to conduct a thorough survey in-field, not only to collect more data on relevant isoglossic features, but also to evaluate how native speakers perceive of subdialectal variation. There are also a number of linguistic phenomena of which I have provided either only a brief or incomplete account; for instance, child-directed speech, and how it differs phonologically and morphosyntactically from typical speech (see Section 2.2.1.1). Further fieldwork is also necessary to determine the productivity of the vocative, the absentative, the unspecified *mi-* possessor, verb diminutivisation, and particle reduplication, among others. The internal structure of verbs, particularly concerning the incorporation of novel elements such as loanwords, is also worthy of future study, as is the compositionality of person-number suffixes in the VTA paradigm. Concerning the Woods Cree sound system, the exact acoustic cues differentiating long and short vowels are still not entirely clear, nor are the cues or patterns associated with stress on the sentence level. The nature of the marginal phonemes /l/ and /r/ is also not fully established.

There are also a number of avenues for future development of the various language tools discussed in this thesis. Firstly, I intend to create an online interface for the text corpus using the concordancing tool KORP (Borin et al. 2012), which has been used previously for corpora of Plains Cree. Secondly, I intend to expand both the text corpus and audio corpus, and to link their contents for texts with existing audio recordings (and vice versa). This linkage could also be used to provide audio recordings for entries on *itwîwina*, although eliciting recordings expressly for this purpose would be an equally valid approach. Finally, linking the Woods Cree text corpus to the other corpora which I have constructed for Plains Cree (see Dacanay & Arppe, forthcoming) and Swampy Cree is another long-term goal, given the extensive typological similarities between the languages.

The morphological model also requires further improvement. One obvious path here is the expansion of its lexicon, which may be accommodated through the inclusion of the *Westfall & Castel English-Cree Dictionary*, the English-Cree section of the CCD, or the conversion of entries from existing Plains Cree dictionaries into Woods Cree (with L1 speaker oversight). The

inclusions of a productive model for numerals and common derivational processes (such as deverbal nominalisation) are also long term goals, as is the modelling of internal verb structure.

4.2 Further Applications

Although the intelligent dictionary *itwîwina* is a major initial step in expanding the pool of online resources for Woods Cree, it is far from complete. Firstly, entries on *itwîwina* currently have no example sentences. An obvious solution to this would be to automatically draw sentences containing the relevant entry lexeme from the text corpus; this functionality, however, has yet to be implemented. Entries also currently lack audio recordings. As mentioned, these too could be extracted from the audio corpus; however, for the purposes of clear and consistent audio quality, these might best be obtained through elicitation in the field.

The morphological model can also be used for a variety of other purposes, aside from as a parser and a dictionary aide. The model and its spell-relax rules could be easily adapted into a Woods Cree spellchecker, allowing computer-assisted writing for Woods Cree (at least, in the SRO) in the same way as it has been available for majority languages such as English for decades. The model could also be used to develop independent computer-assisted language learning tools, applications with which learners could test their morphosyntactic awareness in the language using procedurally generated language lessons (i.e. Bontogon (2016)). In any case, the development of a morphological model for Woods Cree is merely a foundational step, opening the door for a variety of other potential avenues of technology development for the language.

4.3 Summary

My overriding objective throughout this thesis has been to increase the quality and quantity of existing documentation and language resources in Woods Cree, both for academics and for Woods Cree communities. I will provide here a broad summary of how I have, at least to a measure, achieved this aim.

In the first Section, I discussed the historical and contemporary geographic distribution of Woods Cree, its current demographic vitality, and the specific challenges it faces in the 21st century. In addition, I provide an overview of its relationship with neighbouring Western Cree dialects, chiefly Plains Cree and Swampy Cree, particularly concerning their level of mutual intelligibility. I then discuss subdialectal variation within Woods Cree, dividing it into Eastern varieties, which tend towards morphosyntactic conservatism, and Western varieties, which tend towards morphosyntactic innovation. After this, I provide an account of the scant existing language resources for Woods Cree, before describing the process by which I synthesised the majority of this existing material into a text corpus, the first of its kind for the language.

In the second Section, I provided a detailed descriptive account of Woods Cree. Firstly, I discussed its orthography, and the variability therein. Then, I provided an overview of Woods

Cree phonology, illustrating common phonological processes, syllable structure, and stress assignment. After this, I discussed morphology, with distinct subsections for nouns, pronouns, particles, and verbs, as well as some minor elements of syntax. My claims throughout this section stem primarily from data from the text corpus, the observations of previous linguists, and the counsel of an L1 consultant. This section also contains frequent comparisons between Woods Cree and its neighbouring dialects, which I make to further illustrate their mutual relation.

In the third Section, I presented the synthesis of the descriptive account from Section 2 into a finite-state transducer-based morphological model. The purpose of this model is to be able to provide automatic morphological analyses for morphologically complex Woods Cree words, as well as to be able to automatically produce such complex forms for any given stem. From a structural perspective, this model draws almost entirely from an existing morphological model of Plains Cree, and reuses many elements from other existing language technology for other polysynthetic languages. The model is also equipped with regular expression based rules to account for common morphophonological processes and spelling variations. As an initial demonstration of the practical function of this model, I have made use of the web infrastructure of an existing intelligent Plains Cree dictionary to create the first online dictionary for Woods Cree. Following this, Section 4 provides a brief overview of avenues for future research, followed by the present summary.

Overall, although it represents only the beginnings of increased online presence for Woods Cree, this thesis nonetheless serves to contribute to the advancement of Woods Cree linguistic documentation from both a theoretical and applied standpoint. It has yielded a cross-dialectal descriptive analysis of Woods Cree, a bilingual text corpus of over sixty-three thousand Cree tokens, a morphological model with limited spellcheck functionality, and an intelligent online dictionary using this model. It is my hope that these tools may each be expanded upon and enhanced by the Cree speakers and linguists of the future, and that they may further the revitalisation of their language to a far greater extent than that which I have humbly presented here.

Bibliography

Ahenakew, F., & Wolfart, H. C. (1983). Productive reduplication in Plains Cree. In W. Cowan (Ed.), *Actes du quatorzième congrès des algonquinistes* (pp. 369–377). Carleton University Press

Alderete, J., & Kochetov, A. (2016). Integrating sound symbolism with core grammar: The case of expressive palatalization. *Language*, *93*(4), 731-766.

Arppe, A., Harrigan, A., Schmirler, K. & Wolvengrey, A. (2018). *A morphologically intelligent online dictionary for Plains Cree*. Presentation conducted at the meeting of Stabilizing Indigenous Languages Symposium (SILS).

Arppe, A., Schmirler, K., Harrigan, A. G., & Wolvengrey, A. (2020). A morphosyntactically tagged corpus of Plains Cree. In M. Macaulay & M. Noodin (Eds.), *Papers of the 49th Algonquian Conference* (pp. 1-19). Michigan State University Press.

Arppe, A., Poulin, J., Harrigan, A., Schmirler, K., Dacanay, D., & Makinaw, R. (2022). *êkosi ê-nêhiyawi-pîkiskwêcik maskwacîsihk – Towards a Spoken Dictionary of Maskwacîs Cree*. Presentation conducted at the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL-5).

Arppe, A., Neitsch, A., Dacanay, D., Poulin, J., Hieber, D., and Harrigan, A. (2023). Finding words that aren't there: Using word embeddings to improve dictionary search for low-resource languages. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)* (pp.144–155). Association for Computational Linguistics.

Bakker, P. (2006). Algonquian verb structure: Plains Cree. *LOT Occasional Series*, *5*(4), 3-27.

Beesley, K. R., & Karttunen, L. (2003). *Finite state morphology*. Center for the Study of Language and Information.

Bennett, J. A., & Berry, J. W. (1991). Cree literacy in the syllabic script. In D.R. Olson & N. Torrance (Eds.) *Literacy and Orality* (pp. 90-104). Cambridge University Press.

Bloomfield, L. (1925). On the Sound System of Central Algonquian. *Language*, *1*(4), 130-156.

Bloomfield, L. (1928). The Plains Cree Language. *Proceedings of the Twenty-Second International Congress of Americanists*. 427-431

Bloomfield, L. (1946). Algonquian. In H. Hoijer (Ed.), *Linguistic Structures of Native America* (pp. 85-129). Viking Fund Publications in Anthropology.

Bontogon, M. (2016). *Evaluating nêhiyawêtân: A computer assisted language learning (CALL) application for Plains Cree* (Honours thesis).

Borin, L., Forsberg, M., Roxendal, J. (2012). Korp – the corpus infrastructure of Språkbanken. In N. Calzolari (Ed.) *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 474-478). European Language Resources Association (ELRA).

Bowers, D., Arppe, A., Lachler, J., Moshagen, S., Trosterud, T. (n.d.). A Morphological Parser for Odawa. , *2nd Workshop on Computational Methods for Endangered Languages*. Association for Computational Linguistics.

Brightman, R. (1985). The Indefinite Possessor Prefix in Woods Cree. *International Journal of American Linguistics*, *51*(4), 353-356.

Brightman, R. (1993). *Grateful Prey: Rock Cree Human-Animal Relationships*. University of California Press.

Brightman, R. (2007). *Ācaðōhkīwina and ācimōwina: traditional narratives of the Rock Cree Indians*. University of Regina Press.

Castel, R. J., & Westfall, D. (2001). *Castel's English-Cree Dictionary and Memoirs of the Elders*. Brandon University Northern Teacher Education Program.

Cenerini, C. A. M. (2014). *Relational Verbs: Paradigm and Practice in a Manitoba dialect of Swampy Cree* (Master's thesis).

Charles, C., Goulet, K., McKenzie, M. (2021). *Colin Charles Cree Dictionary*. Lac La Ronge Indian Band.

Comrie, B., Haspelmath, M., & Bickel, B. (2008). *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses.* Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics of the University of Leipzig.

Cook, C. & Muehlbauer, J. (2010). *A Morpheme Index of Plains Cree*. Unpublished manuscript.

Dacanay, D. & Arppe, A. (forthcoming). misi-mîkiwâhp pêsêkinosa ohci: A Corpus of Miscellaneous Plains Cree Texts. In *Papers of the Fifty-Fifth Algonquian Conference.* Michigan State University Press.

Dacanay, D. & Schmirler, K. (2022). An Analogy-Based Alternative to "Counter-Intuitive" Grammatical Animacy in Plains Cree/nêhiyawêwin. *Fifty-Fourth Algonquian Conference* [Presentation].

Ellis, C. D. (1971, revised 2015). Cree Verb Paradigms. *International Journal of American Linguistics*, *37*(2), 76-95.

Ellis, C. D. (1973). A proposed standard Roman orthography for Cree. *Western Canadian Journal of Anthropology, 3*(4), 1-37.

Ellis C. D. (2000). *Spoken Cree I*. University of Alberta Press

Ellis C. D. (2004). *Spoken Cree II*. University of Alberta Press

Ellis C. D. (2016). *Spoken Cree III*. University of Alberta Press

Ethnologue. (2015). *Woods Cree*. https://www.ethnologue.com/language/cwd

Evert, S., & Hardie, A. (2011). Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011 Conference* (pp. 1-21). University of Birmingham.

Goddard, I. (1990). Primary and Secondary Stem Derivation in Algonquian. *International Journal of American Linguistics*, *56*(4), 449-483.

Goddard, I. (1994). The west-to-east cline in Algonquian dialectology. In W. Cowan (Ed.), *Papers of the Twenty-Fifth Algonquian Conference* (pp. 187-211). Carleton University Press.

Godden, B. (2019). Summary Paradigm Tables (y- and th-dialects). *Cree Literacy Network*. https://creeliteracy.org/2019/06/18/summary-paradigm-tables-the-decoder-ring-of-cree-verbs-y-and-th-dialects/

Gordon, K. (1965). Testing intelligibility among Chippewa Cree dialects. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session*, *9*.

Greensmith, J. M. (1985a). *Phonological Variants in Pukatawagan Woods Cree*. (Master's thesis).

Greensmith, J. M. (1985b). Future Markers in Woods Cree. In W. Cowan (Ed.), *Papers of the Sixteenth Algonquian Conference* (pp. 63-72). Carleton University Press.

Harrigan, A. G., Schmirler, K., Arppe, A., Antonsen, L., Trosterud, T. (2017). Learning from the computational modelling of Plains Cree verbs. *Morphology*, *27*(4), 565-98.

Hewson, J. (1993). *A computer-generated dictionary of Proto-Algonquian*. University of Ottawa Press.

Howse, J. (1844). *A Grammar of the Cree Language, with which is Combined an Analysis of the Chippeway Dialect*. J. G. F. & J Rivington.

Hulden, M. (2009). Foma: a finite-state compiler and library. In *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics* (pp. 29–32). Association for Computational Linguistics

James, D. (1982). Past Tense, Imperfective Aspect, and lrreality in Cree. In W. Cowan (Ed.), *Papers of the Thirteenth Algonquian Conference* (pp. 143-160). Carleton University Press.

James, D. (1984). The Dubitative in Moose Cree. In W. Cowan (Ed.), *Papers of the Fifteenth Algonquian Conference* (pp. 287-302). Carleton University Press.

James, D. (1991). Preterit Forms in Moose Cree as Markers of Tense, Aspect, and Modality. *International Journal of American Linguistics, 57*(3), 281-297.

Jurafsky, D. & Martin, J. H. (2016). *Speech and Language Processing* (pp. 102-33). Prentice-Hall.

Kadlec, D. (2022). *A Computational Model of Blackfoot Noun and Verb Morphology*. (Master's thesis).

Kang, E. (2017). Morphological variation in Cree inflection. *Toronto Working Papers in Linguistics (TWPL)*, *39*.

Karttunen, L., & Beesley, K. R. (2005). Twenty-Five Years of Finite-State Morphology. In Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M, Westerlund, H., & Yli-Jyrä, A. (Eds.), *Inquiries into Words, Constraints and Contexts Festschrift for Kimmo Koskenniemi on his 60th Birthday* (pp. 71-83). CSLI Studies in Computational Linguistics.

Kazeminejad, G., Cowell, A., Hulden, M. (2017). Creating lexical resources for polysynthetic languages—the case of Arapaho. In A. Arppe, J. Good, M. Hulden, J. Lachler, A.

Palmer, L. Schwartz (Eds.), *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 10-18). : Association for Computational Linguistics.

Koskenniemi, K. (1983). *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production.* University of Helsinki. Department of General Linguistics.

Lacombe, A. (1874). *Dictionnaire de la langue des Cris*. C. O. Beauchemin et Valois

LeClaire, N. & Cardinal, G. (2002). *Alberta Elders' Cree Dictionary/alberta ohci kehtehayak nehiyaw otwestamâkewasinahikan*. University of Alberta Press.

Leighton, A. (1982). *Ethnobotany of the Nihīthawăk, Saskatchewan Woods Cree of the TH (đ) Dialect*. (Master's thesis).

Lindén, K., & Pirinen, T. (2009). Weighted finite-state morphological analysis of Finnish compounding with hfst-lexc. In K. Jokinen & E. Beck (Eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)* (Vol. 4, pp. 89–95). Northern European Association for Language Technology.

MacKenzie, M. E. (1980). *Towards a Dialectology of Cree-Montagnais Naskapi.* Doctoral dissertation, University of Toronto Press

MacKenzie, M. (1992). Negative Markers in East Cree and Montagnais. In W. Cowan (Ed.), *Papers of the Twenty-Third Algonquian Conference* (pp. 274-284). Carleton University Press.

Melnychuk, T. (2003). *Diminutive Consonant Harmony in Several Dialects of Cree*. (Master's thesis).

Mii Dash Geget. (2023). *Proto-Algonquian*, Mii Dash Geget. https://miidashgeget.wordpress.com/2019/01/26/proto-algonquian/

Mikolov, T., Sutskever, I., Kai Chen, Corrado, G. & Dean, J. (1990). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26(2), 3111–3119.

Moeller, S., Kazeminejad, G., Cowell, A., Hulden, M. (2018). A Neural Morphological Analyzer for Arapaho Verbs Learned from a Finite State Transducer. In J. L. Klavans (Ed.), *Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages* (pp. 12-20). Association for Computational Linguistics.

Muehlbauer, J. (2007). Evidence for three distinct nominal classes in Plains Cree. *Natural Language Semantics*, 15(2), 167-186.

Murdoch, J. S. (1981). *Syllabics: A Successful Educational Innovation.* (Master's thesis).

NWT Literacy Council (2022). *Languages of the Land: A Resource Manual for Aboriginal Language Activists*. Office of the Languages Commissioner of the Northwest Territories.

Okimāsis, J., & Wolvengrey, A. (2008). *How to Spell it in Cree (The Standard Roman Orthography)*. Miywāsin Ink.

Oxford, W. (2017). Algonquian grammar myths. *Toronto Working Papers in Linguistics (TWPL)*, *39*.

Oxford, W. (2023a). *Consonant inventories from Proto-Algonquian to the daughter languages*. Manuscript, University of Manitoba. http://home.cc.umanitoba.ca/~oxfordwr/algling/consonantchanges.html.

Oxford, W. (2023b). *Concise Proto-Algonquian verb paradigms*. Manuscript, University of Manitoba. http://home.cc.umanitoba.ca/~oxfordwr/algling/consonantchanges.html.

Pentland, D. (1974). Diminutive Consonant Symbolism in Algonquian. In W. Cowan (Ed.), *Papers of the Sixth Algonquian Conference* (pp. 237-252). National Museum of Canada.

Pentland, D. (1977). *néhiyawasinahikéwin: A Standard Orthography for the Cree Language*. Saskatchewan Indian Federated College.

Pentland, D. (1978). A Historical Overview of Cree Dialects. In W. Cowan (Ed.), *Papers of the Ninth Algonquian Conference* (pp. 104-126). Carleton University Press.

Pentland, D. (2003). The Missinipi Dialect of Cree. In H. C. Wolfart (Ed.), *Papers of the Thirty-Fourth Algonquian Conference* (pp. 287-301). University of Manitoba Press.

Pirinen, F. A., Trosterud, T., Moshagen, S., Wiechetek, L. (2023). GiellaLT: an infrastructure for rule-based language technology tool development. *Journal for the Digital Humanities in Southern Africa*, *4*(1), 1-7.

Pirinen, F. A., Moshagen, S., Hiovain-Asikainen, K. (2023). GiellaLT — a stable infrastructure for Nordic minority languages and beyond. In T. Alumäe & M. Fishel (Eds.), *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)* (pp. 643-649). : University of Tartu Library.

Ratt, S. (2014). *Woods Cree Stories*. University of Regina Press.

Ratt, S. (2016). *Mâci-nêhiyawêwin / Beginning Cree*. University of Regina Press.

Ratt, S. (2023). *kâ-pî-isi-kiskisiyân / The Way I Remember*. University of Regina Press.

Reinholz, C. (2002). On the Characterization of the Cree Question Particle. In H. C. Wolfart (Ed.), *Papers of the Thirty-Third Algonquian Conference* (pp. 378-391). : University of Manitoba Press.

Rock Cree Language Council. (2023). A Sociolinguistics Survey of the Peter Ballantyne Cree Nation in Northern Saskatchewan (2023). *Rock Cree Language Council*. https://nihithawiwin.ca/

Russell, D. R. (1990). *The 18th Century Western Cree and their Neighbours: Identity and Territory*. (Master's thesis).

Schmirler, K. (2016). Acoustic Characteristics of Preaspirated Stops in Plains Cree. *Canadian Acoustics*, 44(3), 142-3.

Schmirler, K. & Arnhold, A. (2017). Acoustic correlates of stress in Maskwacîs Plains Cree. *Fiftieth Algonquian Conference* [Presentation]. https://drive.google.com/file/d/1zvlzMCq3yJ6SwdvQr40OXu9kjXrDEz37/view

Schmirler, K., Arppe, A., Trosterud, T., Antonsen, L. (2018). Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments. In N. Calzolari et al. (Eds.) *Proceedings*

of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA).

Smith, J. G. (1981). Western Woods Cree. In J. Helm (Ed.), *Handbook of North American Indians* (Vol. 6, pp. 256-270). Smithsonian Institution.

Smith, J. G. (1987). The Western Woods Cree: Anthropological Myth and Historical Reality. *American Ethnologist*, *14*(3), 434-448.

Snoek, C., Thunder, D., Lõo, K., Arppe, A., Lachler, J., Moshagen, S., Trosterud, T. (2014). Modeling the Noun Morphology of Plains Cree. In J. Good, J. Hirschberg, O. Rambow (Eds.), *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 34-42). Association for Computational Linguistics.

Spence, G. (1996). *Omushkego Cree Syllabic Final Report*. Queen's Printer for Ontario.

Starks, D. J. (1987). An Introduction to Woods Cree: The Case of South Indian Lake. In W. Cowan (Ed.), *Papers of the Eighteenth Algonquian Conference* (pp. 335-344). Carleton University Press.

Starks, D. J. (1992). *Aspects of Woods Cree Syntax*. Doctoral dissertation, University of Manitoba Press

Starks, D. J., & Ballard, E. (2003). A Short Note on Changes in Pre-Aspiration in Woods Cree. *Linguistica Atlantica*, 24, 1-7.

Starks, D. J., & Ballard, E. (2005). Woods Cree /ð/: An Unusual Type of Sonorant. *International Journal of American Linguistics*, 71(1), 102-115.

Statistics Canada. (2021a). *Mother tongue by geography*, 2021 Census. Retrieved January 4, 2024 from Statistics Canada: https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/dv-vd/language-langue/index-en.html

Statistics Canada. (2021b). *Mother tongue by language spoken most often at home, other language(s) spoken regularly at home, knowledge of official languages and first official language spoken: Canada, provinces and territories, census metropolitan areas and census agglomerations with parts*, 2021 Census. Retrieved January 4, 2024 from Statistics Canada: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810018301

Swan, I. (2000). *Language Shift: A Study of Three Generations within a Cree Family* (Master's thesis).

Torkornoo, D., Junker, M., Owen, C. (2021). *Woodland Cree Conversation* (*Version 1.1*) [Mobile app]. App Store. https://apps.apple.com/ca/app/woodland-cree-conversation/id1581282722

UNESCO. (2010). *Atlas of the World's Languages in Danger* (Map 2). UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000187026

Van Eijk, J., & Collette, V. (2020). Plains Cree animacy–inanimacy hierarchy. *Studies in Language*, 45(4), 840-864.

Watkins, E. (1865). *A Dictionary of the Cree Language as Spoke by the Indians of the Hudson Bay Company's Territories*. Society for Promoting Christian Knowledge.

Wolfart, H. C. (1973). Plains Cree: A Grammatical Study. *Transactions of the American Philosophical Society*, 63(5), 1-90.

Wolfart, H. C., & Ahenakew, F. (2000). *They Knew Both Sides of Medicine: Cree Tales of Curing and Cursing Told by Alice Ahenakew*. University of Manitoba Press.

Wolvengrey, A. E. (1998). On the Spelling and Pronunciation of First Nations Languages and Names in Saskatchewan. In P. C. Douaud (Ed.), *Prairie Forum* (Vol. 23, pp. 113-125). Canadian Plains Research Center.

Wolvengrey, A. E. (2006). Prospective Aspect in the Western Dialects of Cree. *International Journal of American Linguistics*, *72*(3), 397-407.

Wolvengrey, A. E. (2011a). Semantic and pragmatic functions in Plains Cree syntax. Doctoral dissertation, Universiteit van Amsterdam

Wolvengrey, A. E. (2011b). *Cree: Words / nêhiyawêwin: itwêwina*. University of Regina Press.

Wolvengrey, A. E. (2013). iyaw! Whoever Talks about (Cree) Interjections. *Forty-Ninth Algonquian Conference* [Presentation]

Wolvengrey, A. E. (2021). Plains Cree Personal Names. *Anthropological Linguistics*, 63(1-2), 127-151.

Wolvengrey, A. E. (2022). *Plains Cree Grammar*. https://plainscree.algonquianlanguages.ca/grammar/

Yamamoto, M. (1999). *Animacy and reference: A cognitive approach to corpus linguistics*. John Benjamin's Publishing Company.

Appendices

Appendix A: Nomenclature

Perhaps the most common alternative name for Woods Cree is ‘th-dialect Cree’, a reference to Woods Cree’s conspicuous use of phoneme /ð/ as a reflex of the Proto-Algonquian *r (Goddard 1994). However, historically, the term ‘Woods Cree’ has also been used to refer to dialects lacking this /ð/ reflex. For example, in his 1874 treatise *Grammaire de la langue des cris*, Pr. Albert Lacombe use the term ‘Cris des bois’ to refer both to the (now-extinct) dialect of Cree spoken at Île-à-la-Crosse, a dialect which preserved the Proto-Algonquian *r as /r/, and to dialects with /ð/ (Lacombe 1874:x). In addition to this, the terms ‘Woodlands Cree’, ‘Woodland Cree’, and ‘Bush Cree’ are also in common usage, although all three of these names are also occasionally used to refer to northern dialects of Plains Cree[75]. In particular, ‘Bush Cree’ is frequently used to refer to the dialect of Plains Cree spoken in the South Slave region of the Northwest Territories (NWT Literacy Council 2022). The related Woods Cree term *sakâw-ithiniwak*, which may be variously translated as ‘Bush people’ or ‘Woods people’, is also a common ethno-cultural term of self-reference among the Woods Cree (Leighton 1982:4); its Plains Cree equivalent (*sakâw-iyiniwak*) is used to refer both to Woods Cree people in particular, as well as remote, ‘backwoods’ Cree people in general (Wolvengrey 2011b). A similar term, ‘Thickwoods Cree’, is used by anthropologists to refer to the Cree speaking inhabitants of the northern forested regions of Saskatchewan, Alberta, and British Columbia (e.g. Russell 1990, Brightman 1993); however, these groups primarily speak Northern Plains Cree, not Woods Cree in the linguistic sense. The terms ‘Rock Cree’ and ‘Rocky Cree’ (*asinîskâwi-nîhithawîwin*) are two further, predominantly anthropological terms, being used chiefly to refer to the Woods Cree language as it spoken near the border between Manitoba and Saskatchewan, particularly in and around the Peter Ballantyne Cree Nation. These terms are often used self-referentially by Woods Cree speakers in this area; however, some Northern Plains Cree speakers in Alberta also use this label.

Despite this almost dizzying variety in potential nomenclature, as mentioned in Section 1.2, most fluent speakers seldom feel the need to overtly disambiguate their own dialect within the Cree language as a whole, and, unless prompted otherwise, will generally refer to their language in English simply as ‘Cree’. Endonymically, these speakers typically use the terms *nîhithawîwin* (lit. ‘Cree language’), *nîhithawi-pîkiskwîwin* (lit. ‘Cree language’ or ‘Cree speech’), *ithinîmowin* (lit. ‘the people’s language’), or even simply *nipîkiskwîwininân* (lit. ‘our language’) in reference to Woods Cree; however, none of these terms have yet found common usage in English.

[75] There is, for example, a First Nation near Peace River in northern Alberta called the Woodland Cree First Nation, which, despite its name, consists entirely of Northern Plains Cree speaking communities

Appendix B: Excerpts from Western Cree L1 Speakers Concerning Inter-Dialectal Intelligibility

Speaker 1: Rose Makinaw, Plains Cree L1 speaker from Maskwacîs, AB

> "I've been to Lac La Ronge a couple of times [and] I do understand what they're saying. They use three languages, Cree, English, and French, and I'm fortunate I took some French, so I know what they're saying. I understand them … For me it wasn't difficult [although] I can't speak for anybody else. [My French] helped me, when I went to that community … I spent two weekends at Lac La Ronge, and that family we stayed with, I had no problem conversing with them, and they understood me, even though we spoke mostly all Cree. You kinda have to know when you exchange the 'th' for the 'y', or even an 'n' dialect. You have to think fast to figure out what the word is, but based on the context, you can usually tell what they're saying … The further east I went, when I was in Québec, I had a problem understanding. The odd word popped out, but for a full conversation with them, I had trouble understanding."

From an interview I conducted with Rose on the 5th day of December, 2023.

Speaker 2: Solomon Ratt, Woods Cree L1 speaker from Stanley Mission, SK

> "So my schoolmate told me this, *kâya môhcohkâso*! ['Don't act a fool!']. This is at residential school, he's from Sturgeon Lake [a Plains Cree speaking community in Saskatchewan], and I just thought "What?" [laughs]. This is the first time I've ever heard this word! … *pâh-pitos mâna nipîkiskwânân ôtî kîwîtinohk*, *nawac nîthanân î-nihtâwîyahk!* [We tend to speak differently here in the North, we speak better] [laughs] Just kidding!"
>
> "These people, the r-dialect in Québec, I cannot understand them. Just some words, just a few words. I was making a presentation in Ottawa, and some people from Great Whale River [an East Cree speaking village in Northern Québec] were listening to me because my last name was Ratt. They figured we were related, because their last names were Ratt. But they couldn't understand what I was saying in Cree, I couldn't understand what they were saying in Cree. They spoke no English, only French and Cree, and I spoke no French … We ended up using their interpreter; they spoke French to her, and then she spoke English to me. That's the only way we could talk!"

Taken from a lecture given at the 2018 *nîhithaw-âcathôhkâniwin kapîsiwin* (Cree Storytelling Camp) at Big Stone Lake (*mistasiniy-sâkahikan*), Saskatchewan[76]

Appendix C: Comparative Dialogue in Plains Cree (P), Woods Cree (W), and Swampy Cree (S)

P: *atâwêwikamikohk cî ê-wî-itohtêyan?*
W: *atâwikamikohk cî i-wî-itohtîyin?*
S: *k'ka itohtên nâ atâwêwikamikohk?*
'Are you going to the store?'

Ignoring the difference in word order, which is variable in all three dialects, note the differing form of the question marker (*cî~nâ*) between dialects. Note also the differing morphology on the verb *itohtê-* / *itohtî-* ('to go'); *ê-wî-itohtêyan* would also be grammatical in Swampy Cree, as would *kika-itohtân* be grammatical in Plains Cree and Woods Cree.

P: *môya cêskwa, kî-kisinâw. wîpac niwî-sipwêhtân. kîkway cî kinitawêyihtên?*
W: *namwâc cîskwa, kî-tahkâyâw. wîpac niwî-sipwîhtân. kinitawîthihtîn nâ kîkwây?*
S: *môna êškwâ, tahkâyâpan. âšay wîpac nikakihtohtân. kinatawêntên nâ kêkwân?*
'Not yet, it was cold (earlier). I'm leaving soon. Do you want anything?'

Note here the differing, although cognatic, negation markers used in the phrase for 'not yet' at the start of each phrase. Note also the difference in morphology on the verb meaning 'to be cold'; the Swampy Cree sentence uses the Preterit form *tahkâyâpan* (which is archaic in Plains Cree and Woods Cree), rather than the past tense prefix *kî-* (although the form *kî-tahkâyâw* would also be grammatical in Swampy Cree). The Plains Cree sentence also uses a different verb for 'cold' (*kisinâw*) than Woods and Swampy, although *tahkâyâw* is also still commonly used in Plains Cree. Finally, note the systematic phonological differences in the verbs for 'to want' in the third sentence (all stemming from the Proto-Algonquian *natawe·rent-). The Plains Cree form uses /j/ (⟨y⟩, *nitawêyiht-*), the Woods Cree form uses /ð/ (⟨th⟩, *nitawîthiht-*), and the Swampy Cree form uses /n/ (*natawênt-*).

P: *môya, apisîs piko nitayâwâw sôniyâw. kiya mâka, kîkwây ê-nitawêyihtaman?*
W: *namwâc, apisîs poko nitayâwâw sôniyâw. kîtha mâka, kîkwây kâ-nitawîthihtaman?*
S: *môna, apišîš piko nitayâwâw šôniyân. kîna mâka wîna, kêkwân wâ otinikêyan?*
'No, I only have a little money. But what about you, what are you getting?'

[76] The full lecture may be found here:
https://www.youtube.com/watch?v=ES0YyRE9A3M&ab_channel=linguaenthusiast

Note again the difference in negation particles (*môya, namwâc*, *môna*), as well the palatalisation of the /s/ to /ʃ/ in the Swampy Cree *apišîš* compared to the Plains and Woods Cree *apisîs*. The systematic /j/~/ð/~/n/ shift is again visible on the 2SG pronoun (*kiya~kîtha~kîna*). Note also the focus marker (*wîna*) in the second sentence in Swampy Cree; although not present here, the equivalent focus markers (*wiya* and *wîtha*) could be optionally inserted into the Plains and Woods Cree sentences in the same position. Finally, note the slight formal difference in the Swampy Cree interrogative pronoun (*kêkwân* rather than *kîkwây*), and the different choice of lexeme in the Swampy Cree sentence for the verb 'to get' (*otinikêyan*) (although note also that *ê-otinikêyan* or *î-otinikîyin* would also be grammatical in the Plains Cree and Woods Cree sentences respectively).

P: *kisîpêkinikan êkwa oski-maskisina*
W: *sîpîkinikan ikwa oski-maskisina*
S: *sôp nêsta oški-maskisina*
'Soap and new shoes'

Note here the difference in tolerance towards loanwords between dialects; while Swampy Cree uses an English loanword for 'soap', Plains Cree uses an in-language neologism. Both terms, the loanword (*sôp*) and the neologism ((*ki)sîpêkinikan*) are attested in Woods Cree, although note the reduced phonological form of the neologism in Woods Cree compared to Plains Cree. Note also the difference in coordinating conjunctions (*êkwa~ikwa~nêsta)*.

Appendix D: East-West Isogloss Maps

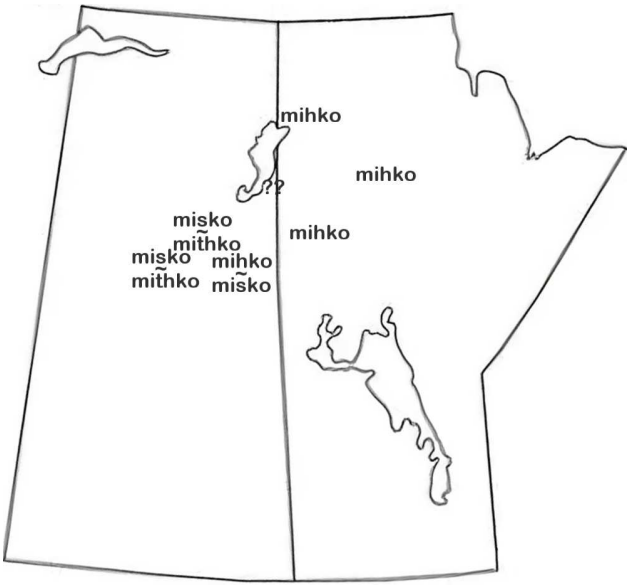

Figure 15. Prevalence of *mithko*, *misko*, and *mihko*; *mihko* predominates in Brochet, South Indian Lake, and Pukatawagan, *mihko* is in free variation with *misko* in Pelican Narrows, and *mithko* and *misko* are in free variation in La Ronge and Stanley Mission. Data is insufficient in Kinoosao to firmly establish a preference

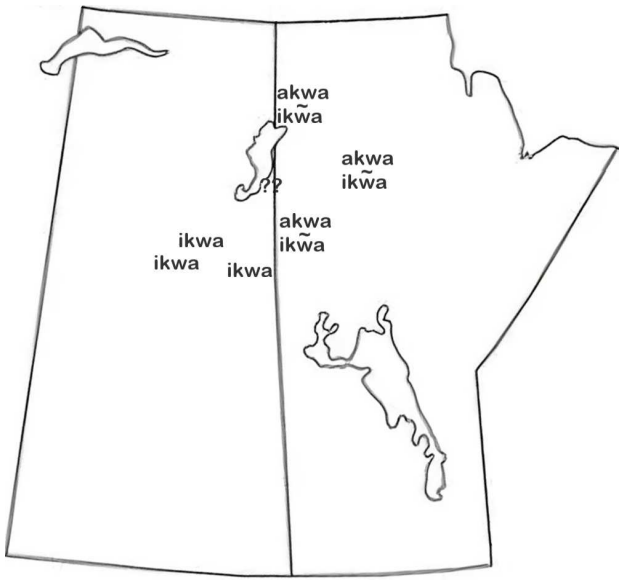

Figure 16. Prevalence of *ikwa* and *akwa*; both are in free variation in Brochet, South Indian Lake, and Pukatawagan, while only *ikwa* is attested in Pelican Narrows, La Ronge, and Stanley Mission. Data is insufficient in Kinoosao to establish a distribution

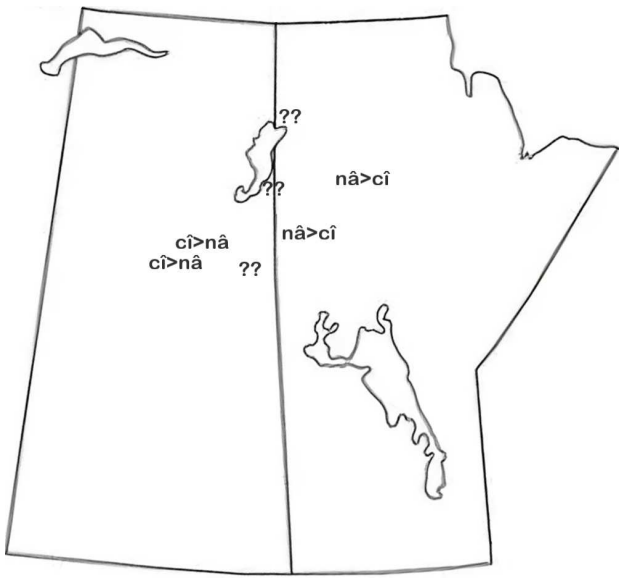

Figure 17. Preference for the closed interrogative markers *nâ* and *cî*; in South Indian Lake and Pukatawagan, *nâ* is preferred over *cî* (*nâ>cî*), whereas in La Ronge and Stanley Mission, *cî* is preferred over *nâ* (*cî>nâ*). Data is insufficient in Brochet, Pelican Narrows, and Kinoosao to establish a preference

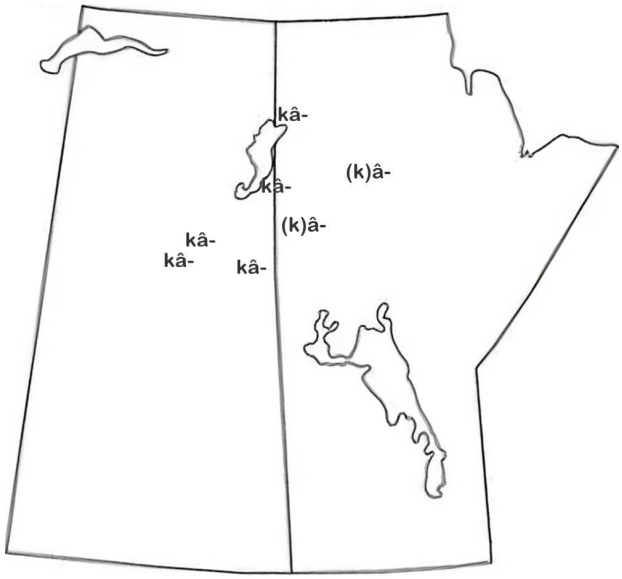

Figure 18. Prevalence of *kâ-* reduction; *kâ-* is regularly reduced to *â-* in South Indian Lake and Pukatawagan, but not in Brochet, Kinoosao, Pelican Narrows, La Ronge, or Stanley Mission

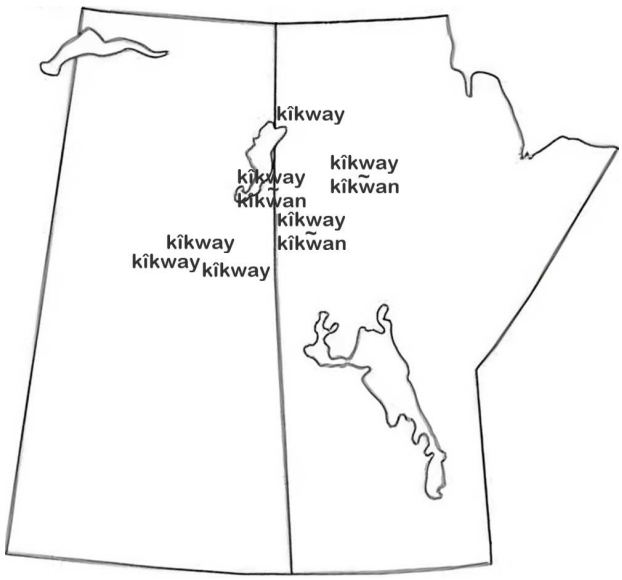

Figure 19. Prevalence of *kîkway* and *kîkwan*; both are attested in South Indian Lake, Pukatawagan, and Kinoosao, but only *kîkway* is attested in Brochet, Pelican Narrows, La Ronge, and Stanley Mission

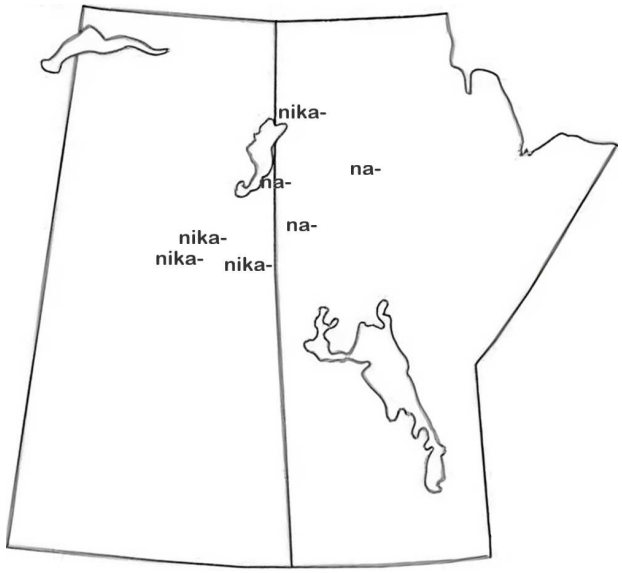

Figure 20. Prevalence of the future marker *na-*; *na-* is preferred in South Indian Lake, Pukatawagan, and Kinoosao, but unattested in Brochet, Pelican Narrows, La Ronge, and Stanley Mission

Appendix E: Woods Cree Phonology (with Common Allophonic Variants)

	Bilabial	Dental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	/p/ ⟨p⟩ [b]		/t/ ⟨t⟩ [d]			/k/ ⟨k⟩ [g] [ɣ] [ʔ]	
Affricate			/t͡s~t͡ʃ/ ⟨c⟩ [d͡ʒ]				
Fricative		/ð/ ⟨th⟩	/s/ ⟨s⟩				/h/ ⟨h⟩
Nasal	/m/ ⟨m⟩		/n/ ⟨n⟩ [ŋ]				
Approximant	/w/ ⟨w⟩				/j/ ⟨y⟩		

Table 32. Consonants of Woods Cree (/IPA/, ⟨SRO⟩, and [allophonic variants])

	Front		Mid		Back	
High	/i/ ⟨i⟩ [ɪ] [e] [ə]	/iː/ ⟨î⟩ [e]				
Mid					/o/ ⟨o⟩ [u] [ʊ] [ə]	/oː/ ⟨ô⟩ [u]
Low			/a/ ⟨a⟩ [ʌ] [e] [ə]	/aː/ ⟨â⟩		

Table 33. Vowels of Woods Cree (/IPA/, ⟨SRO⟩, and [allophonic variants])

Appendix F: The Woods Cree Numeral System

Like other Cree dialects, Woods Cree has a decimal numeral system, and productively forms ordinals using the suffix *-wâw* (although in Eastern varieties, this may also surface as *-wâ*):

	Cardinal	Ordinal
1	*piyak*	*piyakwâ*(*w*)
2	*nîso*	*nîswâ*(*w*)
3	*nisto*	*nistwâ*(*w*)
4	*niyo, nîwo*	*nîwâ*(*w*)
5	*niyânan*	*niyânanwâ*(*w*)
6	*nikotwâsik*	*nikotwâsikwâ*(*w*)
7	*tîpakohp*	*tîpakohpwâ(w)*
8	*ayinânîw*	*ayinânîwâ*(*w*)
9	*kîkâ-mitâtaht* *kîkâc mitâtaht*	*kîkâ-mitâtahtwâ*(*w*) *kîkâc mitâtahtwâ*(*w*)
10	*mitâtaht* *mitâht*	*mitâtahtwâ*(*w*), *mitâhtwâ*(*w*)

Table 34. Cardinal and ordinal numbers 1-10

There is no traditional word for zero; typically, either *ma kîkway* (‘nothing’) or the English word ‘zero’ is used, although a more recent neologism *wâwiyâthisinahikan* (lit. ‘round written thing’) also exists. Woods Cree also represents nine as *kîkâc mitâtaht* (lit. ‘almost ten’), rather than using a descendant of the Proto-Algonquian term for nine *ša·k; this is a feature shared by Plains Cree (which uses *kêkâ-mitâtaht*), but not by Swampy Cree and Moose Cree (which use the PrA-descended *šâkitâtw*).

Words for multiples of ten are formed with some irregularity using the suffix *-mitanaw* and a linking vowel (usually /o/, but sometimes also produced as /i/):

	Cardinal	Ordinal
20	*nîstanaw*	*nîstanawâ*(*w*)
30	*nistomitanaw,* *nistanaw*	*nistomitanawâ*(*w*) *nistanawâ*(*w*)
40	*nîmitanaw*	*nîmitanawâ*(*w*)
50	*niyânanomitanaw* *niyânomitanaw*	*niyânanomitanawâ*(*w*) *niyânomitanawâ*(*w*)
60	*nikotwâsikomitanaw*	*nikotwâsikomitanawâ*(*w*)
70	*tîpakohpomitanaw*	*tîpakohpomitanawâ*(*w*)
80	*ayinânîwomitanaw*	*ayinânîwomitanawâ*(*w*)
90	*kîkâ-mitâtahtomitanaw* *kîkâc mitâtahtomitanaw*	*kîkâ-mitâtahtomitanawâ*(*w*) *kîkâc mitâtahtomitanawâ*(*w*)
100	*mitâtahtomitanaw*	*mitâtahtomitanawâ*(*w*)

Table 35. Cardinal and ordinal numbers 20-100, in intervals of 10

Numbers between multiples of ten are formed by listing the tens digit and then the ones digit followed by the suffix -*osâp*[77]. For numbers between ten and twenty, the tens digit (*mitâtaht*) is often omitted:

233) a. *nîstanaw nistosâp*
nîso-mitanaw nisto-osâp
two-tens three-past
'twenty-three' (UL)

b. *nistanaw niyânosâp*
nisto-mitanaw niyânan-osâp
three-tens five-past
'thirty-five' (PK)

c. *mitâtaht piyakosâp*
mitâtaht piyak-osâp
ten one-past
'eleven' (LR)

d. *piyakosâp*
piyak-osâp
one-past
'eleven' (SM)

Multiples of one hundred are typically formed by adding the ordinal form of the relevant factor before *mitâtahtomitanaw*, however, the unmarked cardinal form is also sometimes used. These generalisations also apply to multiples of one thousand:

[77] Probably derived from the Root/Initial element *sâp-* ('through')

234) a. *piyakwâw kihci-mitâtahtomitanaw*
piyak-wâw kihci-mitâtahto-mitanaw
one-ORD great-ten-tens
'one thousand' (SM)

b. *niyo mitâhtomitanaw*
niyo mitâhto-mitanaw
four ten-tens
'four hundred' (PK)

In practice, most modern Woods Cree speakers, even those who are highly fluent, tend to use English numerals in speech, especially for higher numbers. This is particularly evident in giving calendrical dates and years, which are almost exclusively produced in English:

235) a. sixteen *nitahtwâskîwinân*
sixteen ni-tahtwâskîwinî-n
sixteen 1SG.IND-have.such.age(VAI)-1SG.IND
'I was just sixteen years old' (PK)

b. nineteen-fifty *pâham â-kî-mâcipathik*
nineteen-fifty pâham (k)â-kî-mâcipathi-k
nineteen-fifty perhaps CNJ-PST-start(VII)-3SG.CNJ
'it was perhaps 1950 it started' (PK)

Appendix G: Marginal Paradigms

Example 1: The Relational

Historically, all VAIs and VTIs in Woods Cree were produced in either Relational form or Non-Relational form. Relational form marked the presence or involvement of an additional third-person argument of the verb not specified by the verb's inflection (Cenerini 2014):

236) a. *ya. kahkithaw anihi owîhthowiniwâ nikiskîthihtamwân*
ya kahkithaw anihi o-wîhthowin-iwâ(w)-(a)
yes all DEM.MED.INAN.PL 3PL.POSS-name(NI)-3PL.POSS-PL
ni-kiskîthiht-amwân
1SG.IND-know(VTI)-1SG.REL
'Yes, all of them, I know their name(s).' (PK)

b. *ôta macî iyakwanik anikik â-kanawîthihtamwat anihi ayahâwak*

ôta	macî	iyakwanik	anikik
here	you.see	PRON-ANIM-PL	DEM-MED-ANIM-PL

kâ-kanawîthiht-amwat	anihi	ayhâwak
CNJ-look.after(VTI)-2SG.REL	DEM.MED.ANIM.OBV	PRON.ANIM.PL

'Here, you know, those ones, [in the picture] you keep, those ones' (PK)

Relational forms were using the relational suffix *-w* (or *-amw* for VTI stems), followed by the appropriate person-number inflection expected for the given verb's mode. For Independent verbs, this person-number affixation is identical to Independent verbs elsewhere. For Conjunct and Imperative verbs, a different set of person-number affixes was used; the Swampy Cree and Moose Cree forms of these affixes are detailed in Ellis (2016:159), and the (simpler) Plains Cree forms are recorded in Wolfart (1973:60)

Relational forms are uncommon across contemporary Woods Cree, but are more readily found in the Eastern variety than in the West. Modern speakers generally mark additional participants in verbs by selecting different Finals, such as *-amaw* (which marks ditransitive VTAs), or simply use the typical VAI or VTI form, implying additional participants purely pragmatically.

237) a. *nikî-kakwî-asotamawâw ta-atoskâtamwak otatoskîwin*

ni-kî-kakwî-asotamaw-âw	ta-atoskâtam-wak
1SG.IND-PST-try-promise(VTA)-1SG>3SG.IND	FUT-do.work(VTI)-3PL.IND

o-t-atoskîwin
3SG.POSS-LC-work(NI)
'I also promised that I would do his work for him.' (SM)

b. *niwanikiskisin â-isithihkâsocik anikik*

ni-wanikiskisi-n	(k)â-isithihkâso-cik	anikik
1SG.IND-forget(VAI)-1SG.IND	CNJ-be.called.thus(VAI)-3PL.CNJ	DEM.MED.ANIM.PL

'I forget what their names were.' (PK)

This loss of the Relational is paralleled in Plains Cree, which retains only a handful of fossilised Relational forms (Harrigan et al. 2017:574), and contrasts with Swampy Cree and Moose Cree, both of which still use the Relational productively (Cenerini 2014).

Example 2: The Dubitative

Historically, evidentiality was encoded morphologically in Woods Cree verbs, with a separate Dubitative mode existing for the Independent and Conjunct paradigms of each verb. Howse does

not provide a full paradigm, but provides several examples of what he deems 'Doubtful' verb forms, which closely resemble modern Swampy Cree and Moose Cree dubitative forms:

238) a. Nippá-tookè-nik
nipâtokînik
nipâ-tokînik
sleep(VAI)-3PL.DUB
'they sleep, I suppose' (Howse 1844:261)
Note: Compare *nipâtokwênak* in modern Swampy Cree

b. Tàn etéthemèwunè
tân(a) itîthimiwânî
tân(a) itîthim-iwânî
PRON.INTERR.ANIM.SG think(VTA)-2SG>1SG.DUB
'whatever thou mayest think of me' (Howse 1844:261)
Note: Compare *itênimiwânê* in modern Swampy Cree

The Howse grammar is the only written source to contain a significant number of dubitative forms; however, even this source provides only 9 marked examples. As such, I have been unable to confidently construct a full paradigm from the attested forms; however, I have included a list of all attested dubitative forms from Howse in Appendix H. All of the dubitative affixes present in the Howse grammar can be directly linked to contemporary dubitative affixes in Swampy Cree and Moose Cree (Ellis 2016); I therefore believe it likely that the historical system in Woods Cree closely resembled this extant system in its neighbouring dialects.

Dubitative verb forms have fallen entirely out of use in contemporary Woods Cree, with various evidential particles having displaced it (e.g. *îtokwî* ('probably'), *îsa* ('reportedly')); in the corpus, I have yet to identify a single modern example of the historical Dubitative set. One partial exception to this may be found in South Indian Lake, where the dubitative prefix *wâh-* has been reported (Starks 1987); however, this dubitative appears to have no genealogical relationship to historical morphological dubitative, and may be a modern borrowing from Swampy Cree (see, for instance, the third Swampy Cree example sentence in Appendix C).

239) *wâh-itohtîyâni, ka-wâh-wanisininaw*
wâh-itohtî-yâni kika-wâh-wanisin-inaw
DUB-go(VAI)-1SG.SUBJ 12PL.IND-FUT-DUB-get.lost-12PL.IND
'If I had gone, we would be lost' (SL)

It is not entirely clear when the Woods Cree dubitative paradigm fell out of use, as there is exceedingly little documentation of the language between the Howse grammar in 1844 and the

works of Starks and Greensmith in the 1980s; however, through comparison with neighbouring dialects, a date of extinction some time in the early 20th century appears probable. Plains Cree, which had an almost identical Dubitative verb paradigm in the 19th century (Lacombe 1874), saw its Dubitative morphology become "archaic and rare" by the 1920s (Bloomfield 1928). By contrast, Dubitative morphology is still productive (albeit seemingly declining in use (Cenerini 2014:29)) in Swampy Cree and is both productive and frequent in Moose Cree (James 1984).

240) *âsay maci-kîsikâtokwê walawîtimihk ê-papihtikwêk*
âsay maci-kîsikâ-tokwê walawîtimihk ê-papihtikwê-k
now be.bad.weather(VII)-3SG.DUB outside CNJ-thunder(VII)-3SG.CNJ
'It must be a bad day outside now from the sound of the thunder.' (Moose Cree)
(Ellis 2016:11)

Example 3: The Preterit

Historically, the Independent and Conjunct verb Orders were further subdivided into two tense categories, the Neutral, which provides no information on tense, and the Preterit, which described continuous past actions. Of these two, only the Neutral survives in contemporary Woods Cree, with temporal information now being conveyed chiefly through tense prefixes, and the Neutral tense category having become the default Present tense. The Preterit, by contrast, has fallen mostly out of use; Starks (1992:64) reports that her consultants were "unable to provide complete Preterit paradigms", and that most forms which did occur were produced by older speakers. That said, the Preterit is not entirely extinct in form, and isolated modern instances occur in the corpus across the Eastern and Western communities. However, its meaning has shifted away from the continuous past in most instances; the modern Preterit is chiefly used to denote events which are hypothetical, conditional, or otherwise irrealis, similar in function to the (still productive) Subjunctive:

241) a. *mâka kîspin ka-wanâhtâwiht otihtâwiniwâw nawac ikwa kita-aspihikwak anihi*
mâka kîspin (ki)ka-wanâhtâ-wiht o-t-ihtâwin-iwâw
but if (12PL)- FUT-disturb(VTI)-12PL.PRET 3PL.POSS-LC-habitat-3PL.POSS
nawac ikwa kita-aspih-ikwak anihi
more and FUT-treat.badly-3`-3PL.IND DEM.MED.ANIM.OBV
'Disturbance to their habitat may give predators a hunting advantage' (UL)

b. *nika-kî-nipahikohtay askitipathiwin*
ni-ka_kî-nipah-iko-htay askitipathiwin
1SG-can-kill(VTA)-INAN-1SG.PRET perforated.ulcer
'I could have died from a perforated ulcer' (SM)

In contemporary Woods Cree, Preterit forms have only survived in the Independent Order, even though Conjunct Preterit verbs are recorded in Howse (1844:301). This is perhaps due to the temporally unbound nature of the Conjunct, precluding the need for highly marked tense specifiers such as the Preterit; however, I concede that I cannot fully explain this idiosyncrasy. Several detailed sets of paradigms for the Preterit in Swampy Cree and Moose Cree may be found in Ellis (2016), as may a historical paradigm set for the Plains Cree Preterit be found in Wolfart (1973:43); the historical Woods Cree set appears to have been essentially identical to those found in its neighbours, barring regularised sound shifts

Although no longer productive, modern occurrences of the Preterit do not appear to be systematically fossilised; that is, there do not appear to be any particular verb stems which are more prone to taking Preterit morphology than others. Orthographic variation and homography with other, more productive affixes makes ascertaining an exact number of Preterit verbs in the corpus difficult; however, of the several dozen instances which have been identified, no single lexeme occurs in the Preterit multiple times across multiple speakers.

The Preterit is often coordinated with the modal prefixes *ka-kî-* ('can, may') and *ta-kî-* ('should, would'), presumably to accentuate its irrealis semantics. Most identified instances of the Preterit in the corpus were coordinated in this manner. For example:

242) a. *kika-kî-sâkihisohtay kâ-isi-sâkihitân*
ki-ka_kî-sâkihiso-htay kâ-isi-sâkih-itân
2SG.IND-can-love.self(VAI)-2SG.PRET CNJ-thus-love(VTA)-1SG>2SG.CNJ
'you would love yourself as I love you' (SM)

b. *akwâni kîsta ta-kî-manâskwîhtay nâ âta witha?*
akwâni kîsta ta_kî-manâskwî-htay nâ âta witha
then 2SG.ADD.FOC would-take.up.arms-2SG.PRET INTERR nevertheless
'Then you would have taken up arms, too, right?' (PK)

Like the Dubitative, it is unclear when the Preterit began to fall out of use, with the most precise estimate from the textual record being at some point between 1844 and the 1980s. However, it is perhaps illuminating that, while vestigal Preterit forms still occasionally occur in contemporary speech, vestigal Dubitative forms are entirely absent. This may suggest either that the Dubitative fell out of use earlier than the Preterit, or that the Dubitative was markedly less frequent than the Preterit to begin with. The truth is likely a combination of both; Howse, for instance, records much less documentation on the Dubitative than he does the Preterit (suggesting, perhaps, that he had either encountered it less frequently or considered it less significant), and in neighbouring Plains Cree, Bloomfield describes the Preterit as "archaicizing" by the 1920s (1928:429),

contrasting with his description of the Dubitative in the same work, which he deems already "rare and archaic".

Appendix H: Dubitative Forms Provided by Howse (1844:261) with Modern Swampy Cree Equivalents

243) Nippátookènik
nipâtokînak
nipâ-tokînak
sleep(VAI)-3PL.IND.DUB
'they sleep, I suppose'
Note: Compare Swampy Cree *nipâtokwênak*

244) I`eskootáytookènik
ayîskohtîtokînak
ayîskohtî-tokînak
tire.from.walking(VAI)-3PL.IND.DUB
'they are tired (with walking), I suppose'
Note: Compare Swampy Cree *ayîskohtêtokwênak*

245) Sáhkeháytookènik
sâkihîtokînak
sâkih-îtokînak
love(VTA)-3PL>3`.IND.DUB
'they love them, I suppose'
Note: Compare Swampy Cree *sâkihîtokwênak*

246) Sáhkehikóotookènik
sâkihikotokînak
sâkih-ikotokînak
love(VTA)-3`>3PL.IND.DUB
'they are loved by them, I suppose'
Note: Compare Swampy Cree *sâkihikotokwênak*

247) Kuckwáychemik kutta itóotaywákwè
kakwîcimik kita-itohtîwâkwî
kakwîcim-ik kita-itohtî-wâkwî
ask(VTA)-2SG>3PL.IMP.IMM FUT-go(VAI)-3PL.CNJ.DUB
'ask (thou) them if they go or not'
Note: Compare Swampy Cree *kakwêcimik kita-itohtîwâkwê*

248) Númmă ne kiskétheten ittè gà atháywákwè
nama nikiskîthihtîn ita kâ-athahâwâkwî
nama nikiskîthihtîn ita kâ-athahâwâkwî
NEG 1SG.IND-know(VTI)-1SG.IND there CNJ-bury(VTA)-3PL>3`.CNJ.DUB
'I do not know (it) the place where they may have laid him'
Note: Compare Swampy Cree *nama nikiskênihtên ita kâ-anahâwâkwê*

249) Keespin sákehéwunè
kîspin sâkihiwânî
kîspin sâkih-iwânî
if love(VTA)-2SG>1SG.CNJ.DUB
'if thou love me'
Note: Compare Swampy Cree *kîspin sâkihiwânê*

250) Tàn etéthemèwunè
tân itîthimiwânî
tân itîthim-iwânî
INTERR think(VTA)-2SG>1SG.CNJ.DUB
'whatever thou mayest think (of) me'
Note: Compare Swampy Cree *tân itênimiwânê*

251) Keespin úntowéthemewáigwè
kîspin nitawîthimiwîkwî
kîspin nitawîthim-iwîkwî
if seek(VTA)-2PL>1SG.CNJ.DUB
'if ye seek me'
Note: Compare Swampy Cree *kîspin nitawênimiwêkwê*

Appendix I: Demonstration of Relative Affix Orderings for Verbs

252) a. <u>Tense>Manner/Direction>Reduplication>Lexical Preverb>Stem>Person/Number</u>
kî-ohci-mâh-manâ(y)-itwîw
kî-ohci-mâh-manâ-itwî-w
PST-from-RDPLS-respect-say(VAI)-3SG.IND
's/he cared about what s/he said' (PK)

b. <u>Aspect>Manner/Direction>Stem>Person/Number</u>
mîkwâ-nitawi-pîskiskwîstamâkîci
mîkwâ-nitawi-pîskiskwîstamâkî-ci
simultaneously-go-testify(VAI)-3SG.SUBJ
's/he may go to testify at the same time' (UL)

c. Person>Tense>Aspect>Lexical Preverb-Stem-Person/Number
nika-ati-mitho-pimohtân
ni-ka-ati-mitho-pimohtî-n
1SG.IND-FUT-start-good-walk(VAI)-1SG.IND
'I will start to walk a good path' (SM)

d. Person>Modal>Manner/Direction>Stem>Person/Number
ninohtî-pakaski-nihithawân
ni-nohtî-pakaski-nihithawî-n
1SG.IND-want-fluently-speak.Cree(VAI)-1SG.IND
'I want to speak Cree fluently' (Southend, SK)

e. Conjunct>Tense>Manner/Direction>Manner/Direction>Lexical Preverb>Stem>Person/Number
kâ-kî-pî-isi-nîhithawi-pimâtisiyân
kâ-kî-pî-isi-nîhithawi-pimâtisi-yân
CNJ-PST-come-thus-Cree-live(VAI)-1SG.CNJ
'I lived a Cree way of life' (SM)

Appendix J: Example of a Simplified FST-Based Morphological Model for Afrikaans

Afrikaans verbs have only three inflected forms; the past participle (achieved with the prefix *ge-*), the present participle (achieved with the suffix *-ende*), and the infinitive (which is morphologically null. Thus, for the verb stem *ly* ('suffer'), there are three possible inflected forms:

ly - 'to suffer'
gely - 'suffered'
lyende - 'suffering'

An FST model for Afrikaans verbs would have three possible inputs: `PST+` (which has the output of adding *ge-* to the left edge of the stem (e.g. *gely* ('suffered')), `+PRP` (which has the output of adding *-ende* to the right edge of the stem (e.g. *lyende* ('suffering')), and `+INF` (which does not modify the surface form of the output stem (e.g. *ly* ('to suffer')). Using LEXC, FSTs are structured into discrete slots known as 'lexica', with each lexicon containing a related set of potential inputs. When an initial state is passed through the FST, it passes through each of these lexica sequentially. An example of a LEXC-based Afrikaans verb FST is provided below:

```
LEXICON Prefix
PST+:ge Roots;
```

```
Roots;

LEXICON Roots
ly Suffixes;

LEXICON Suffix
+PRP:ende #;
+INF:0 #;
#;
```

Line by line, the above FST reads as follows. Firstly, in the `Prefix` lexicon, the only possible prefix is `ge`. If this prefix is selected, the underlying value of `PST+` is added, and the surface value of `ge` is added to the lexeme in question. A colon separates the underlying analysis (`PST+`) from the surface realisation (`ge`). The next lexicon destination of the lexeme (`Roots`) is separated from these analyses with a space. Alternatively, the model may proceed directly to the `Roots` lexicon if no underlying `PST+` is present. Within the `Roots` lexicon, there is only one possible root, `ly`, which must be selected and cannot be skipped. Once this has been selected, the model proceeds to the `Suffix` lexicon. There are three possible options in the `Suffix` lexicon; the model may add the underlying value `+PRP` (insodoing adding the surface value of `ende` to the right edge), add the underlying value of `+INF` (which adds nothing to the surface representation) or skip the `Suffix` lexicon entirely. The model then reaches the end (as indicated by the character #), and outputs the final wordform.

In addition to being able to produce all possible inflected forms of *ly*, this FST would also be able to reconstruct an input form of *ly* (with morphological tags) based solely on an inflected surface form. For instance, it would be able to take the form `gely` and extrapolate that the outputs necessary to create that form were `ge` (associated with the tag `PST+` in the `Prefix` lexicon) and `ly` (which is listed in the `Roots` lexicon), and therefore that the input necessary to create this form must have been `PST+ly`. In this way, even this simple Afrikaans FST is simultaneously able to act as both an analyser and generator of inflected forms. For a description of how this is accomplished from a computational perspective, as well as for much more detailed description of finite state modelling in general, see Beesley & Karttunen (2003)

The results of the above model may be further improved using post-hoc rewrite rules, sets of regular-expression-based rules applied after the relevant affixation process(es) intended to simulate the morphophonological processes which occur between the underlying and surface forms. For example, in Afrikaans, when the present participle ending -*ende* is added to stems ending in /f/ (⟨f⟩), the /f/ is voiced to /v/ (⟨w⟩). Without rewrite rules, the current Afrikaans model would produce the present participle forms of *f*-final stems incorrectly; for instance, for the

participle of verb *sterf* (‘to die’), it would produce the form **sterfende* instead of the correct form *sterwende* (‘dying’). This could be remedied using a rewrite rule, namely;

```
f -> w || _ e n d e;
```

This rule reads as follows: `f` must become `w` when it immediately precedes the character sequence `ende`. This could be improved further by the addition of explicitly indicated boundary markers to each morpheme, to prevent the rule from erroneously applying in contexts where an `f` precedes the sequence `ende`, but is not a participle (e.g. *Mafende* (a surname)). This is demonstrated in the following set of lexica and rewrite rules, using the characters < and > for prefix and suffix boundaries respectively (alongside the escape character %, as < and > also have alternative syntactic functions in LEXC). In this set, the `f -> w` rule only occurs when it precedes a morpheme boundary marker and then `ende`, preventing application in non-participles:

```
LEXICON Prefix
PST+:ge%< Roots;
Roots;

LEXICON Roots
ly Suffixes;
sterf Suffixes;

LEXICON Suffix
+PRP:%>ende #;
+INF:%>0 #;
#;

! Rewrite Rule
f -> w || _ %> e n d e;
```

While these markers solve the issue of representing this particular morphophonological process, they also result in the model producing output forms which contain morpheme boundary characters (e.g. *sterw>ende* instead of *sterwende*). As such, we may add a second rewrite rule, applying after the first, which deletes all morpheme boundary characters as the last step of the model, after the other rewrite rule(s) has already made use of them. The final model would thus be as follows:

```
! Finite State Model
LEXICON Prefix
```

```
PST+:ge%< Roots;
Roots;

LEXICON Roots
ly Suffixes;
sterf Suffixes;

LEXICON Suffix
+PRP:%>ende #;
+INF:%>0 #;
#;

! Rewrite Rules
f -> w || _ %> e n d e;
%> -> 0;
```

Appendix K: Example .yaml Files for Nouns

NA (e.g. *kwâpahikan* (‘ladle’))

Noun+N+A+Sg:Output

```
kwâpahikan+N+A+Sg:kwâpahikan
```

Noun+N+A+Pl:Output

```
kwâpahikan+N+A+Pl:kwâpahikanak
```

Noun+N+A+Obv:Output

```
kwâpahikan+N+A+Obv:kwâpahikana
```

Noun+N+A+Distr:Output[78]

```
kwâpahikan+N+A+Distr:kwâpahikaninâhk
```

Noun+N+A+Der/Dim+Sg:Output

```
kwâpahikan+N+A+Der/Dim+Sg:kwâpahikanis
```

etc. (with plural, obviative, and distributive forms)

NI (e.g. *cîmân* (‘canoe’))

Noun+N+I+Sg:Output

```
cîmân+N+I+Sg:cîmân
```

Noun+N+I+Pl:Output

```
cîmân+N+I+Pl:cîmâna
```

Noun+N+I+Loc:Output

[78] Only included if the given noun is tagged in the lexicon as a noun which may take the distributive

cîmân+N+I+Loc:cîmânihk

Noun+N+I+Der/Dim+Sg:Output
cîmân+N+I+Der/Dim+Sg:cîmânis
etc. (with plural, and locative forms)

NA
Noun+N+A+Px1Sg+Sg:Output
kwâpahikanN+A+Px1Sg+Sg:nikwâpahikan
Noun+N+A+Px1Sg+Pl:Output
kwâpahikanN+A+Px1Sg+Sg:nikwâpahikanak
Noun+N+A+Px1Sg+Obv:Output
kwâpahikanN+A+Px1Sg+Sg:nikwâpahikana
Noun+N+A+Px1Sg+Distr:Output
kwâpahikanN+A+Px1Sg+Sg:nikwâpahikaninâhk
etc. (with 2SG, 3SG, 1PL, 12PL, 2PL, 3PL, and 3` possessed forms and diminutives)

NI
Noun+N+I+Px1Sg+Sg:Output
cîmân+N+I+Px1Sg+Sg:nicîmân
Noun+N+I+Px1Sg+Pl:Output
cîmân+N+I+Px1Sg+Pl:nicîmânak
Noun+N+I+Px1Sg+Loc:Output
cîmân+N+I+Px1Sg+Loc:nicîmânihk
etc.

Appendix L: Example .yaml Files for Verb Classes

Example 1: VII .yaml files

Being the least inflectionally complex verbal subclass, the VII .yaml files were correspondingly the simplest among the verbs, testing if the model could correctly produce the 3SG, 3PL, 4SG, and 4PL forms of a given VII stem in the Independent, the Conjunct, and the Subjunctive, as well as testing if these forms could be produced in each of the preset tense-Order-mood combinations (except for the imperative, as VIIs lack these). In the .yaml files of certain VII subclasses, I also tested if the model would recognise subdialectal variation in certain forms, such as those of the Conjunct plural. This made a total of 55 form-analysis pairs for VII-2n and 60 for VII-2v. I provide the following demonstration for the the VII-2v stem *miskwâ-* (‘to be red’); for brevity’s sake, I provide only the forms for the Present Independent and Conjunct, as well as the Subjunctive/Future-Conditional:

Independent Present

```
miskwâw+V+II+Ind+3Sg: miskwâw
miskwâw+V+II+Ind+4Sg: miskwâthiw
miskwâw+V+II+Ind+3Pl: miskwâwa
miskwâw+V+II+Ind+4Pl: miskwâthiwa
```

etc.

Conjunct Present

```
PV/î+miskwâw+V+II+Cnj+3Sg: î-miskwâk
PV/î+miskwâw+V+II+Cnj+4Sg: î-miskwâthik
PV/î+miskwâw+V+II+Cnj+3Pl: î-miskwâki
PV/î+miskwâw+V+II+Cnj+3Pl+Var/East: [î-miskwâkwâ,
                                     î-miskwâkiht,
                                     î-miskwâhki]
PV/î+miskwâw+V+II+Cnj+4Pl+Var/East: î-miskwâthikwâ
```

etc.

Future Conditional

```
miskwâw+V+II+Fut+Cond+3Sg: miskwâki
miskwâw+V+II+Fut+Cond+4Sg: miskwâthiki
miskwâw+V+II+Fut+Cond+3Pl: miskwâkwâwi
miskwâw+V+II+Fut+Cond+4Pl: miskwâthikwâwi
```

For the four VII stems for which .yaml files were created (*miskwâ-*, *mâthâtan-*, *mîthwâsin-*, and *pimamon-*), the model was able to attain a perfect score for each, being able to both recognise and generate all of the anticipated inflected forms manually recorded in the .yaml files.

Example 2: VAI .yaml Files

For VAIs, I constructed the .yaml files to test if the model could correctly produce the 1SG, 2SG, 3SG, 1PL, 12PL, 2PL, 3PL, 3`, and X (Indefinite Actor) forms of a given VAI stem in the Independent, Conjunct, and Subjunctive (in all tense-Order-mood combinations), as well as if it could correctly produce the 2SG, 12PL, and 2PL forms in the Delayed and Immediate Imperative. In total, this made 126 form-analysis pairs in each .yaml file:

Independent Present

```
atoskîw+V+AI+Ind+1Sg: nitatoskân
atoskîw+V+AI+Ind+2Sg: kitatoskân
atoskîw+V+AI+Ind+3Sg: atoskîw
atoskîw+V+AI+Ind+1Pl: nitatoskânân
atoskîw+V+AI+Ind+12Pl: [kitatoskânaw, kitatoskânânaw]
atoskîw+V+AI+Ind+2Pl: kitatoskânâwâw
atoskîw+V+AI+Ind+3Pl: atoskîwak
```

```
atoskîw+V+AI+Ind+4Sg/Pl: atoskîthiwa
atoskîw+V+AI+Ind+X: [atoskâniwiw, atoskâniwan]
```

etc.

Conjunct Present

```
PV/î+atoskîw+V+AI+Ind+1Sg: î-atoskîyân
PV/î+atoskîw+V+AI+Ind+2Sg: [î-atoskîyan, î-atoskîyin]
PV/î+atoskîw+V+AI+Ind+3Sg: î-atoskît
PV/î+atoskîw+V+AI+Ind+1Pl: î-atoskîyâhk
PV/î+atoskîw+V+AI+Ind+12Pl: î-atoskîyahk
PV/î+atoskîw+V+AI+Ind+2Pl: î-atoskîyîk
PV/î+atoskîw+V+AI+Ind+3Pl: [î-atoskîcik, î-atoskîtwâw]
PV/î+atoskîw+V+AI+Ind+4Sg/Pl: î-atoskîthit
PV/î+atoskîw+V+AI+Ind+X: [î-atoskîhk, î-atoskâniwik,
                          î-atoskâniwahk][79]
```

etc.

Future Conditional

```
atoskîw+V+AI+Ind+1Sg: atoskîyâni
atoskîw+V+AI+Ind+2Sg: [atoskîyani, atoskîyini]
atoskîw+V+AI+Ind+3Sg: atoskîci
atoskîw+V+AI+Ind+1Pl: atoskîyâhki
atoskîw+V+AI+Ind+12Pl: atoskîyahki
atoskîw+V+AI+Ind+2Pl: atoskîyîko
atoskîw+V+AI+Ind+3Pl: atoskîtwâwi
atoskîw+V+AI+Ind+4Sg/Pl: atoskîthici
atoskîw+V+AI+Ind+X: [atoskîhki, atoskîhkwâwi]
```

Imperative

Immediate

```
atoskîw+V+AI+Imp+Imm+2Sg: atoskî
atoskîw+V+AI+Imp+Imm+12Pl: atoskîtân
atoskîw+V+AI+Imp+Imm+2Pl: atoskîk
```

Delayed

```
atoskîw+V+AI+Imp+Del+2Sg: atoskîhkan
atoskîw+V+AI+Imp+Del+12Pl: atoskîhkahk
atoskîw+V+AI+Imp+Del+2Pl: atoskîhkîk
```

[79] Strictly speaking, it may be more accurate to consider Indefinite Actor forms of VAIs to be distinct VII stems, making this a process of derivation, rather than inflection. However, the Plains Cree model already possessed the code to recognise these forms, hence my inclusion of them here

On account of the substantial formal differences between certain inflected wordforms in the VAI-2 paradigm in the Eastern and Western varieties, I included tests in the VAI-2 .yaml file to determine if the model could account for these differences. This increased the number of form-analysis pairs to 161 in the VAI-2 .yaml file:

Independent Present

```
pimisin+V+AI+Ind+1Sg: nipimisinin
pimisin+V+AI+Ind+1Sg+Var/East: nipimisin
pimisin+V+AI+Ind+2Sg: kipimisinin
pimisin+V+AI+Ind+2Sg+Var/East: kipimisin
pimisin+V+AI+Ind+3Sg: pimisin
pimisin+V+AI+Ind+1Pl: nipimisinân
pimisin+V+AI+Ind+12Pl: [kitpimisinaw, kipimisinânaw]
pimisin+V+AI+Ind+2Pl: kipimisinâwâw
pimisin+V+AI+Ind+3Pl: pimisinwak
pimisin+V+AI+Ind+4Sg/Pl: pimisinithiwa
pimisin+V+AI+Ind+X: [pimisinâniwiw, pimisinâniwan]
```

etc.

Conjunct Present

```
PV/î+pimisin+V+AI+Cnj+1Sg: î-pimisiniyân
PV/î+pimisin+V+AI+Cnj+1Sg+Var/East: î-pimisinân
PV/î+pimisin+V+AI+Cnj+2Sg: [î-pimisiniyan,
                            î-pimisiniyin]
PV/î+pimisin+V+AI+Cnj+2Sg+Var/East: î-pimisinan
PV/î+pimisin+V+AI+Cnj+3Sg: î-pimisihk
PV/î+pimisin+V+AI+Cnj+1Pl: î-pimisiniyâhk
PV/î+pimisin+V+AI+Cnj+1Pl+Var/East: î-pimisinâhk
PV/î+pimisin+V+AI+Cnj+12Pl: î-pimisiniyahk
PV/î+pimisin+V+AI+Cnj+12Pl+Var/East: î-pimisinahk
PV/î+pimisin+V+AI+Cnj+2Pl: î-pimisiniyîk
PV/î+pimisin+V+AI+Cnj+2Pl+Var/East: î-pimisinîk
PV/î+pimisin+V+AI+Cnj+3Pl: î-pimisihkwâw
PV/î+pimisin+V+AI+Cnj+4Sg/Pl: î-pimisinithit
PV/î+pimisin+V+AI+Cnj+X: [î-pimisinâniwik,
                          î-pimisinâniwahk,
                          î-pimisinihk]
```

etc.

Future Conditional

```
pimisin+V+AI+Fut+Cond+1Sg: pimisiniyâni
pimisin+V+AI+Fut+Cond+1Sg: pimisinâni
```

```
pimisin+V+AI+Fut+Cond+2Sg: [pimisiniyani,
                            pimisiniyini]
pimisin+V+AI+Fut+Cond+2Sg+Var/East: pimisinani
pimisin+V+AI+Fut+Cond+3Sg: [pimisinici, pimisihki]
pimisin+V+AI+Fut+Cond+1Pl: pimisiniyâhki
pimisin+V+AI+Fut+Cond+1Pl: pimisinâhki
pimisin+V+AI+Fut+Cond+12Pl: pimisiniyahki
pimisin+V+AI+Fut+Cond+12Pl+Var/East: pimisinahki
pimisin+V+AI+Fut+Cond+2Pl: pimisiniyîko
pimisin+V+AI+Fut+Cond+2Pl+Var/East: pimisinîko
pimisin+V+AI+Fut+Cond+3Pl: [pimisinitwâwi,
                            pimisihkwâwi]
pimisin+V+AI+Fut+Cond+4Sg/Pl: pimisinithici
pimisin+V+AI+Fut+Cond+X: [pimisihki, pimisihkwâwi]
```

Imperative

Immediate

```
pimisin+V+AI+Imp+Imm+2Sg: pimisini
pimisin+V+AI+Imp+Imm+12Pl: pimisinitân
pimisin+V+AI+Imp+Imm+2Pl: pimisinik
```

Delayed

```
pimisin+V+AI+Imp+Del+2Sg: pimisinîhkan
pimisin+V+AI+Imp+Del+12Pl: pimisinîhkahk
pimisin+V+AI+Imp+Del+2Pl: pimisinîhkîk
```

For each of the six VAI stems for which .yaml files were created (*atoskî-*, *mâto-*, *mîciso-*, *nipâ-*, *nîhithawî-*, and *pimisin-*), the verb model attained a perfect score in generation and recognition.

Example 3: VTI .yaml Files

Although the expected output forms differ (at least for VTI-1), the inflectional categories included in the VTI .yaml files were identical to those in the VAI .yaml files, with a total of 90 form-analysis pairs included for VTI-1 and 126 each for VTI-2 and VTI-3:

Independent Present

```
nâtam+V+TI+Ind+1Sg: ninâtîn
nâtam+V+TI+Ind+2Sg: kinâtîn
nâtam+V+TI+Ind+3Sg: nâtam
nâtam+V+TI+Ind+1Pl: ninâtînân
nâtam+V+TI+Ind+12Pl: [kinâtînaw, kinâtînânaw]
nâtam+V+TI+Ind+2Pl: kinâtînâwâw
nâtam+V+TI+Ind+3Pl: nâtamwak
```

```
nâtam+V+TI+Ind+4Sg/Pl: nâtamithiwa
```

Conjunct Present

```
PV/î+nâtam+V+TI+Cnj+1Sg: î-nâtamân
PV/î+nâtam+V+TI+Cnj+2Sg: î-nâtaman
PV/î+nâtam+V+TI+Cnj+3Sg: î-nâtahk
PV/î+nâtam+V+TI+Cnj+1Pl: î-nâtamâhk
PV/î+nâtam+V+TI+Cnj+12Pl: î-nâtamahk
PV/î+nâtam+V+TI+Cnj+2Pl: î-nâtamîk
PV/î+nâtam+V+TI+Cnj+3Pl: î-nâtahkwâw
PV/î+nâtam+V+TI+Cnj+4Sg/Pl: î-nâtamithit
```

Future Conditional

```
nâtam+V+TI+Fut+Cond+1Sg: nâtamâni
nâtam+V+TI+Fut+Cond+2Sg: nâtamani
nâtam+V+TI+Fut+Cond+3Sg: nâtahki
nâtam+V+TI+Fut+Cond+1Pl: nâtamâhki
nâtam+V+TI+Fut+Cond+12Pl: nâtamahki
nâtam+V+TI+Fut+Cond+2Pl: nâtamîko
nâtam+V+TI+Fut+Cond+3Pl: nâtahkwâwi
nâtam+V+TI+Fut+Cond+4Sg/Pl: nâtamithici
```

Imperative

Immediate

```
nâtam+V+TI+Imp+Imm+2Sg: nâta
nâtam+V+TI+Imp+Imm+12Pl: nâtîtân
nâtam+V+TI+Imp+Imm+2Pl: nâtamok
```

Delayed

```
nâtam+V+TI+Imp+Del+2Sg: nâtamohkan
nâtam+V+TI+Imp+Del+12Pl: nâtamohkahk
nâtam+V+TI+Imp+Del+2Pl: nâtamohkîk
```

For each of the three VTI stems for which .yaml files were created, the verb model attained a perfect score in generation and recognition.

Example 4: VTA .yaml Files

For the VTA .yaml files, I included all possible person-number combinations of subject and object (1SG, 2SG, 3SG, 1PL, 12PL, 2PL, 3PL, 3` (or `+4Sg/Pl`), 3`` (or `+5Sg/Pl`), X, and INAN (or `+0Sg/Pl`)) in each tense-Order-mood combination. Additionally, all the paradigms for all five VTA subclasses vary substantially between the Eastern and Western varieties, particularly forms in the Mixed set involving 1PL and 12PL actors; as such, multiple tagged dialectal versions were provided for each relevant form. Predictably, the VTA .yaml files were the largest by far,

containing 699 forms each. In light of this, rather than embedding the contents of such a file in the text, I will provide a link to a representative example as a footnote[80]. However, despite their size, the basic format of the VTA .yaml files is identical to those of the previous.

Like the other subclasses, when the verb model was tested against the five manually created VTA .yaml files (*wîcih*-, *nitonaw*-, *kîskisw*-, *nakat*-, *it*-), it attained a perfect score for each in recognition and generation.

[80]https://github.com/giellalt/lang-cwd/blob/main/test/src/gt-norm-yamls/V-TA-1-w%C3%AEcih%C3%AEw_gt-norm.yaml