A Quantitative Account of Nêhiyawêwin Order: Using mixed-effects modeling to uncover syntactic, semantic, and morphological motivations in Nêhiyawêwin

by

Atticus G. Harrigan

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

Doctor of Philosophy

Department of Linguistics
University of Alberta

University of Alberta

Library Release Form

Name of Author: Atticus G. Harrigan

Title of Thesis: A Quantitative Account of Nêhiyawêwin Order: Using mixed-effects modeling to uncover syntactic, semantic, and morphological motivations in Nêhiyawêwin

Degree: Doctor of Philosophy

Year this Degree Granted: 2023

. . . . . .

Atticus G. Harrigan
Assiniboia 4-32
University of Alberta
Edmonton, AB
Canada, T6G 2G6

Date: . . . . . .

University of Alberta

A Quantitative Account of Nêhiyawêwin Order: Using mixed-effects modeling to uncover syntactic, semantic, and morphological motivations in Nêhiyawêwin

by

Atticus G. Harrigan

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Linguistics

Department of Linguistics

Edmonton, Alberta
Fall 2023

Abstract

This dissertation investigates the underpinnings of the phenomenon of Order in Nêhiyawêwin (Plains Cree) using quantitative methods and the Ahenakew-Wolfart Corpus (Arppe et al., 2020). Instantiated as person-marking allomorphy on the verb, Order is central to verb morphology in Algonquian languages. According to Bloomfield (1946, 97), Algonquian languages have mutually exclusive paradigms within each verb class that serve a number of purposes including marking various moods. Orders do not cleanly map one-to-one onto other grammatical functions, but the system can be thought of as a set of morphological templates. Unlike Semitic languages where morphological templates are attributes of a verb (i.e., each word has one template), Order is a set of templates wherein each verb can alternate. This dissertations approaches Order in Nêhiyawêwin as an alternation between multiple forms that are motivated by morphosemantic features. Importantly, this dissertation explicitly defines Order as those forms traditional referred to as Independent and Conjunct; that is, it does not consider the Imperative to be an instantiation of Order. This allows for an analysis of three main types of alternation at varying levels of granularity which is done through quantitative methodologies. The primary method used for analysis is that of logistic regression (Bresnan et al. (2007), Divjak (2010), and Arppe (2008)). Specifically, mixed effects logistics regression is used, allowing for the analysis of multiple variables as they cooccur. The results of this analysis indicate that, contrary to expectations (and the results of previous research in other languages, such as Arppe (2008), Abdulrahim (2013),

and Divjak and Arppe (2013)), morphosemantic features explained a relatively small amount of variance in each alternation. Instead, it appears that higher level linguistic information, such as discourse planning and reference are more important factors. These results comport with those of Cook (2014). In addition to the study of alternation, this dissertation also presents a set of exemplars that drawn from a corpus and are predicted to be the most likely (or prototypical) forms of each outcome in each alternation (cf. Divjak and Arppe (2013) who used a similar methodology). These example sentences are given in hopes that language learners and educators may use them to identify characteristics of prototypical forms of Order.

Preface

Some of the work in this dissertation has been published (or accepted for publication). Chapter 2 has been adjusted for publication in a future issue of Linguistics Vanguard. This paper was co-written with Dr. Antti Arppe. For this paper, my contribution was in drafting the article, identifying the research questions, analyzing how Order operates in Nêhiyawêwin, and jointly identifying that Order represents a new type of alternation that should be properly described. My co-author was responsible for providing analysis of alternation and recognizing the need for a new type of alternation.

Earlier versions of Chapter 3 were published as Harrigan, A. and Arppe, A. (2021). Leveraging English Word Embeddings for Semi-Automatic Semantic Classification in Nêhiyawêwin (Plains Cree). In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 113–121. Association for Computational Linguistics; as well as Harrigan, A. G. and Arppe, A. (2023). Leveraging Majority Language Resources for Plains Cree Semantic Classification. *In Papers of the Fifty-Second Algonquian Conference*, pages 129–146. Michigan State University Press. In each of these, my co-author Dr. Antti Arppe suggested the use of pre-trained English word vectors on the systemic definitions of Nêhiyawêwin words from Wolvengrey (2001); my contributions were in writing the article, identifying the issue, drafting the code to process the word vectors and create the sentence vectors, performing manual post-processing, developing an evaluation criteria for the final product, and comparing purely automatic and semi-manual processes.

Contents

List of Figures

List of Tables

Glossing Abbreviations

Gloss	Description
ACTOR	Actor of a verb
GOAL	Goal of a verb
NA	Animate Noun
NI	Inanimate Noun
VII	Intransitive inanimate verb
VAI	Animate intransitive verb
VTI	Intransitive inanimate verb
VTA	Transitive inanimate verb
IC	Initial Change
IND	Independent Order
CNJ	Conjunct Order
1	First person
2	Second person
21PL	First person inclusive
3	Third person
3′	Obviative person
POSS	Possesed
SG	Singular
PL	Plural

SG/PL	Unspecified number
DIR	Direct
INV	Inverse
PROX	Proximate
OBV	Obviative
PST	Past tense
DIST	Distal demonstrative
DEM	Demonstrative
FOC	Focus
RFLX	Reflexive
FUT	Future
VOL	Volitional
CON	Continuous
COND	Conditional
NEG	Negative
FUROBV	Further Obviative
Q	Question marker
RDPLW	Weak/light reduplication
RDPLS	Strong/heavy reduplication
UNSPEC	Unspecified actor
THM	theme

Modelling Abbreviations

Gloss	Description
actor.1	First person actor
actor.2	Second person actor
actor.3	Third person actor
actor.obv	Obviative actor
AI.cooking	AI verb of cooking
AI.do	AI verb of action
AI.health	AI verb of health
AI.pray	AI verb of praying
AI.reflexive	Reflexive AI verb
AI.speech	AI verb of speech
AI.state	Stative AI verb
D.actor	Dependent noun actor
D.goal	Dependent noun goal
Dem.actor	Demonstrative actor
Dem.goal	Demonstrative goal
Der.Dim.goal	Diminutive goal
goal.1	First person goal
goal.2	Second person goal
goal.3	Third person goal

goal.obv	Obviative goal
II.natural.land	II verb of land or nature
II.sense	II sensory verb
Med.actor	Medial Demonstrative actor
Med.goal	Medial demonstrative goal
NA.beast.of.burden.actor	Animate beast of burden noun actor
NA.food.actor	Animate food noun actor
NA.persons.actor	Animate human actor noun actor
NA.persons.goal	Animate human actor noun goal
NDA.Relations.actor	Animate dependent noun of kinship actor
NDI.Body.goal	Inanimate body part noun goal
NI.natural.force.goal	Inanimate force of nature noun goal
NI.nominal.goal	Inanimate noun derived from a verb goal
NI.object.actor	Inanimate object noun actor
NI.object.goal	Inanimate object noun goal
NI.place.goal	Place as an inanimate noun goal
Obv.actor	Obviative actor
Pers.actor	Personal pronoun actor
Pl.actor	Plural actor
Pl.goal	Plural goal
Pron.actor	Pronoun actor
Pron.goal	Pronoun goal
Prox.actor	Proximate actor
Prox.goal	Proximate goal
PV.Discourse	Discourse preverb
PV.Move	Movement preverb
PV.Position	Position preverb

PV.Qual	Quality preverb
PV.StartFin	Preverb of starting/finishing
PV.Time	Preverb of time
PV.WantCan	Preverb of desire
Px1Sg.actor	Actor possessed by a singular first person
Px1Sg.goal	Goal possessed by a singular first person
Px3Pl.goal	Goal possessed by a plural third person
Px3Sg.goal	Goal possessed by a singual third person
RdplW	Weak/light reduplication
Sg.actor	Singular actor
Sg.goal	Singular goal
TA.cognitive	TA verb of cognition
TA.do	Ta verb of action
TA.food	TA verb relating to food
TA.money.count	TA verb of money
TA.speech	TA verb of speech
TI.cognitive	TI verb of cognition
TI.do	TI verb of action
TI.money.count	TI verb of money
TI.speech	TI verb of speech

Corpus Abbreviations

Gloss	Description
CMBK-1-2	cmbk-ch1-vers2
CMBK-2-2	cmbk-ch2-vers2
CMBK-3-2	cmbk-ch3-vers2
CMBK-4-2	cmbk-ch4-vers2
CMBK-5-2	cmbk-ch5-vers2
C2GB	C2GB-ARC
AL	AL-RL-C.FIN
SW	SW-CFIN2.D91
VDC2	VDC2-RES.782
EM	EM-CREE5.N97
C7MW	C7MW-ARC.792
JK	JK-C4ARC.798
C4MF	C4MF-ARC.792
AA	AA-c2rev.899
C6IC	C6IC-ARC.792
C8GB	C8GB-ARC.792

Chapter 1

Introduction

This dissertation explores the Nêhiyawêwin phenomenon, Order.[1] Nêhiyawêwin, like all Algonquian languages, is a polysynthetic language with a rich morphological system. The most striking system in Algonquian verbal morphology, apart from perhaps the language's hierarchical/direct-inverse alignment (Wolvengrey, 2011, 53), is the system of Order. Order is instantiated on verbs through a system of allomorphy of the polypersonal argument morphs. While other Algonquian languages differ in their number of Orders, Nêhiyawêwin has three recognized Orders: the Independent, the Conjunct, and the Imperative.

The function of Order has been discussed in the literature, most comprehensively by Wolfart (1973) and Cook (2014). The latter resource is most comprehensive, though it generally focuses only on the difference between the Independent and the Conjunct Orders, placing aside the Imperative Order. This is, in my opinion, valid, but unmotivated in Cook (2014, 11) who justifies her decision as 'There is a third paradigm: the imperative order. The imperative order cannot host most agreement, any of the elements on the far left edge, or most of the preverbs. I will not discuss it further.' Put simply, Cook

[1]Nêhiyawêwin has also been also be referred to as Plains Cree or y-dialect Cree. *Nêhiyawêwin* is the endonym for the language. I have chosen to use this term in this dissertation on the request of multiple native speakers. Their requests are generally motivated by a desire to not use a name given to their language by settlers.

(2014) proposes that the phenomenon of Order is one of clause typing, specifically in the difference between Indexical (not having a prior referrent) and Anaphoric (having a prior referent) clauses. This conclusion was come to after careful hand-analysis of a Nêhiyawêwin corpus. Although I agree with many of the conclusions put forth by Cook (2014), the orientation of this research is decidedly theoretical.

This dissertation will approach the purpose and function on Nêhiyawêwin from a systematic and empirical perspective. Using a corpus (including, in part, all of the texts used by Cook (2014)) and modern computational techniques, this dissertation attempts to uncover the motivations in the choice to use one Order over another. This research is undertaken through the lens of *alternation*. The primary method of analysis this dissertation relies on is mixed-effects logistic regression, based on and building upon the work of Arppe (2008); Bresnan et al. (2007); Divjak (2010); Klavan (2020). By framing Order as a system of alternation, mixed-effects logistic regression allows for the creation of a predictive model, where each of the predictor variables can be evaluated for their effect on the outcome of the alternation. Three types of alternations are investigated: Independent vs. Conjunct (the most straightforward alternation in terms of previous description of Order), Independent vs. ê-Conjunct (the most straightforward alternation in terms of near-synonymy (Cruse, 2000, 157-159)), and the alternation of the various Conjunct types (a more straightforwardly semantic alternation).

Chapter 2 provides a background on Nêhiyawêwin, Order, and the use of alternation in linguistic investigation. This chapter also provides a detailed discussion regarding the nature of Order as an alternation, how this outlook can be used to study the phenomenon, and a detailed justification for ignoring the Imperative mood beyond methodological opportunism.

Next, Chapter 3 presents a study in semi-automatically clustering verbs together for the purposes of predictor generation for the logistic modelling at the centre of this dissertation. This chapter focuses on how one can use pre-existing majority language data

to bootstrap the creation of an ontology for lemmas in a minority language, Nêhiyawêwin. The result of this research, a semantic class for every verb in a dictionary (Wolvengrey, 2001), was used as the main semantic effects in the main statistical modeling of this dissertation.

Following the chapter of semantic classification, Chapter 4 describes and justifies the particular methodologies in statistical modelling. This chapter also details the morphosyntactically tagged corpus that is being used and the ways in which this corpus has been construed as a data set. The main research questions driving this analysis are:

1. Can mixed effects modelling be used in investigating complex morphological phenomenon using a small but richly tagged corpus?

2. Are the alternations between the Independent and Conjunct, the Independent and the ê-Conjunct, and the Conjunct types similarly possible to be modelled, or are some of these alternations more difficult to model than others?

3. What are the variables that increase the likelihood that a verb lemma will occur in a particular Order/outcome?

Chapter 5 presents the results the statistical modelling in three stages: univariate, bivariate, and multivariate. As the latter is of primary interest for this dissertation, it is discussed most in depth. The following chapter, Chapter 6, discusses in detail the multivariate results. This includes not only a discussion of what this means in the general sense of Order as well as how the results frame each outcome, but also how well the statistical modelling performed and what this overall success or failure can tell us about alternations and Order more generally. Finally, Chapter 7 provides a conclusion to this dissertation.

Code used for the analyses presented in this dissertation is available publicly.[2] Although the code functions, the underlying corpus that is analyzed is not able to shared

[2]https://github.com/atticusha/DissertationCode

publicly. Should researchers desire access to the corpus source files, they can contact Dr. Antti Arppe at the University of Alberta. A searchable web-based version of the corpus is also available.[3]

[3]https://korp.altlab.app/

Chapter 2

Background

Nêhiyawêwin (ᓀᐦᐃᔨᐧᐃᐣ ISO 639-3: CRK, also known as Plains Cree) is a member of the Algonquian language family, one of the major language families of North America, spoken across much of Canada and the eastern United States. Nêhiyawêwin is part of two major dialect continua: the Cree-Montagnais-Naskapi languages, which stretch mainly from Alberta to Quebec and Labrador; and the central Canadian Ojibwe-Cree continuum, which occur in central Canada and the Northern United States. [1]

Nêhiyawêwin is the westernmost member of the Cree-Montagnais-Naskapi continuum and is mostly spoken in Alberta, Saskatchewan, and northern Montana. There is said to be approximately 34,000 speakers of Nêhiyawêwin (Ethnologue, 2016), most over the age of 30. This number is likely overestimated, though a previous account by Ethnologue was dubious, with a number of roughly 150. Statistics Canada (2016) reports

[1]The languages/dialects of Cree-Montagnais-Naskapi continuum share many similarities, as can be seen in the following words for 'person', where the reflexes of Proto-Algonquian **l* are given in boldface. The dialects are given in roughly west to east geographical distribution:

(1) Plains: *i**y**iniw*
Woods: *i**th**iniw*
Swampy: *i**n**iniw*
Moose: *i**l**iliw*
Atikamekw: *i**r**iniw*
Ojibwe: inini East: *i**y**iyiw/i**y**iyû/i**y**inû*
Naskapi: *i**y**iyû*
Innu: *i**l**nu/i**n**nu*

3,655 native speakers of 'Plains Cree,' though this number may be higher if respondents reported their native language as 'Cree,' rather than 'Plains Cree.' Wolfart (1973) estimated 20,000 speakers, though the number has likely dropped since then. Although any of these numbers is dwarfed by the number of speakers of majority languages in Canada, Nêhiyawêwin retains a strong presence, particularly for an Indigenous North American language, holding a classification of 5 (Developing) on the Extended Graded Intergenerational Disruption Scale (EGIDS) (Ethnologue, 2016), a system for assessing language vitality based on domains of use, intergenerational transmission, and other sociolinguistic factors (Lewis and Simons, 2012). With its comparatively large speaker base, Nêhiyawêwin has garnered attention from a variety of Americanists, in the form of grammars (e.g. Wolfart 1973, Dahlstrom 2014, Wolvengrey 2011), textbooks (e.g. Okimāsis 2018; Ratt 2016) and an online electronic dictionary (itwêwina[2]).

2.1 Nouns

Nêhiyawêwin exhibits a number of morphosyntactic features that differ considerably from the well-known characteristics of often discussed Indo-European languages. Unlike sex-based gender systems such as those found in many contemporary romance languages, Algonquian languages have a two-way gender or noun classification system contrasting Inanimate with Animate nouns; this grammatical animacy has some basis in semantic animacy: all humans, animals, and trees are Animate. This distinction is not clear-cut though, as *êmihkwân*, 'spoon', *sîwinikan*, 'sugar', and *sêhkêpayîs*, 'automobile' are Animate,[3] and thus the system is considered one of grammatical classification. Notably there are few-to-no examples of clearly alive animals that are grammatically Inanimate. Animacy is relevant to nominal and verbal morphology in Nêhiyawêwin

[2]https:/itwewina.altlab.app

[3]It is worth noting that animacy is not always consistent across dialects of Nêhiyawêwin, or even communities of Nêhiyawêwin. Some words, such as *sîwinikan* 'sugar', are Animate in some dialects and Inanimate in others.

in various ways. Among nouns, this animacy distinction is manifested in two distinct plural markers, {-ak} for Animate and {-a} for Inanimate nouns; archaic singular marking is seen for monosyllabic roots, for example *maskw-a* 'bear (ANIM)' and *wâw-i* 'egg (INANIM)'. Nêhiyawêwin has no grammatical case system, but it does have locative marking, generally *-ihk* for Inanimate nouns (Wolfart, 1973, 1996), with human/animal Animate nouns often not being locativized.

Nêhiyawêwin is a head-marking language, and so the person and number of the possessor is marked on the possessum. Singular possessors are marked only with prefixes: {ni-} for first person, {ki-} for second person, and {o-} for third person. For plural possessors, circumfixes are used: the prefixes are the same as for singular persons, which are matched with a set of suffixes: {ni- -(i)nân} for first person plural exclusive ('ours but not yours'), {ki- -(i)naw} for first person plural inclusive ('mine/ours and yours'), {ki- -(i)wâw} for second person plural ('yours but not ours'), and {o- -(i)wâw} for third person plural. Nêhiyawêwin also distinguishes between alienable and inalienable nouns; the latter category must occur with possession and includes kinship terms and body parts as well as some other intimate possessions or relationships, such as *nôhkom* 'my grandmother' versus **ôhkom* 'grandmother' (Wolfart, 1973, 1996; Wolvengrey, 2011). Some nouns, particularly body parts, are inalienable and may be possessed by a general possessor, as in *mitâs*, '(someone's) pants'.

Within Animate nouns, a pragmatic distinction is made regarding the topicality of a noun when used in the third person. All Animate nouns can occur as either proximate third person (more topical entity in a discourse) and the obviative third person (less topical entity or entities in the discourse). This distinction occurs any time more than one Animate third person occurs in a discourse, such as when one third person Animate entity acts on another or when a third person Animate entity possesses another, as in (1). An obviative Animate noun is marked with the obviative suffix {-a} and no number distinction is made; this is conventionally marked with 3′ (or as the '4th person', with no

number distinction; in this dissertation it will be indicated by OBV in glosses). The further obviative, which occurs when two obviative entities occur in one discourse, necessitating the demotion of one of them, is by convention marked with 3′′ (or as the '5th person', also with no number distinction; in this dissertation it will be indicated by FUROBV). As obviation is based in topicality rather than syntactic roles, it is generally not considered a marker of case. This is further exemplified with respect to verbal constructions below.

(1) atim nâpêw-a tahkwam-ê-w
dog.PROX man-OBV bite-DIR.THM-3SG.OBV

'the (proximate) dog bites the (obviative) man.'

2.2 Verbs

Cree verbs are traditionally classified according to both their transitivity and the animacy of their arguments/participants. There are two classes of intransitive verbs, which can occur with one Inanimate participant (VII—Verb Inanimate Intransitive) or one Animate participant (VAI—Verb Animate Intransitive). The former includes impersonal verbs such as weather terms and stative verbs used attributively to describe Inanimate objects, and the latter includes intransitive actions and attributive verbs used to describe Animate objects (Bloomfield, 1946; Okimāsis, 2018; Wolfart, 1973, 1996). The VII and VAI classes are exemplified in (2) and (3) respectively.

(2) VII

a. wâpiskâ-w
be.white-3SG

'it is white'

b. astotin wâpiskâ-w
hat be.white-3SG

'the hat (Inanimate) is white'

(3) VAI

a. wâpiskisi-w
be.white-3SG

's/he (animate) is white'

b. mîciso-w
eat-3SG

's/he eats, has a meal'

Similarly, there are two classes of transitive verbs, though these are distinguished by the animacy of their second participant, often considered the object: transitive Inanimate verbs (VTI) with an Animate subject and an Inanimate object, and transitive Animate verbs (VTA) with two Animate arguments.[4] Examples are given in (4) and (5); note that there are three different verbs for 'eat' depending on the transitivity and the animacy of participants.

(4) VTI

mîci-w
eat-3SG

's/he eats it (Inanimate)'

(5) VTA

mow-ê-w
eat-DIR.THM-3SG.ACTOR.OBVGOAL

's/he eats it/him (animate)'

As noted above, Nêhiyawêwin does not have a case system to determine syntactic roles. Nouns exhibit obviation, a system in which non-focal, Animate, third persons are

[4]*Subjects* and *objects* are conventionally called *actors* and *goals* in Algonquian literature (Bloomfield, 1946; Wolvengrey, 2011). *Actors* here refer to the do-er of an action or subject of a description, despite the syntactic or semantic role. Similarly, *goals* are any entity that receives a transitive action, regardless of the semantic or syntactic role (e.g. patient, recipient, benefactive, etc.). For this dissertation, I make use of these terms.

marked (Bloomfield, 1946, 94). Together with the directionality system, discussed below, semantic roles are determined through relationships between items rather than simple case marking.

Verbs agree with arguments according to animacy: Inanimate actors for VII and Animate actors for VAI, VTI, and VTA. The Inanimate participant in a clause containing a VTI is the goal of the verb, or some other oblique argument, but not the actor. The person marking on VII, VAI, and VTI verbs corresponds to the person and number of the actor. However, in VTAs, both arguments are Animate and realized in the verbal morphology, with their respective roles determined by obviation and direction morphology, discussed below. Essentially, verbs and their arguments can be thought of as constructions where certain verb stems license a certain number of arguments of particular animacy.

To determine the roles of participants in VTA clauses, Algonquian languages make use of a direct-inverse system (Jacques and Antonov, 2014; Wolfart, 1973). VTAs occur with two Animate participants and there is no grammatical case or fixed word order by which to determine the semantic roles. Instead, direction is used as a method of determining which argument is the actor and which is the goal. In Nêhiyawêwin, direction is determined by the relative topicality of participants, extended beyond the proximate-obviative distinction into a full hierarchy known as the Algonquian person hierarchy, given in (6) (Jolley, 1983). Direction is indicated by a theme morpheme, which indicates that the action is either *direct* or *inverse*. When a more topical participant acts on a less topical participant, the morphology or theme sign is direct (-*â*-, -*ê*-, -*i*-). When the opposite occurs, the morphology or theme sign is inverse (-*ik*(*w/o*)-, -*iti*-). As visualized in (6), second person is ranked topically above first person, and both of these speech act participants are ranked above all third or unspecified[5] persons, wherein obviation applies. Due to this hierarchy, first person acting on second necessarily always occurs with inverse morphology. In this way, these are not passive forms, but

[5]In Nêhiyawêwin, the Unspecified Actor is an actor on a verb where the exact person and number of the actor is not specified. It may be translated as a sort of agentless passive (Wolvengrey, 2011).

simply the only way of indicating first person acting on second. For this and a variety of other reasons not discussed herein, Nêhiyawêwin inverse forms are not considered equivalent to passive voice in languages such as English (Dahlstrom, 2014; Wolfart, 1973; Wolvengrey, 2011).

(6) $2 > 1 >$ Unspecified Actor $> 3 > 3' > 3''$

With obviation marked on both nouns and verbs, sentences such as those in (7)a. are possible in Nêhiyawêwin. Additionally, both obviative and further obviative marking may be needed, depending on the number of third persons lexically specified, as in (7)b. However, when a Nêhiyawêwin VTI is involved, and so there is an Inanimate goal rather than an Animate one, no goal or obviative marking occurs on either the verb, or the Inanimate noun, as in (8) (Wolfart, 1973; Wolvengrey, 2011).

(7) VTA

a. cân pahkwêsikan-a mow-ê-w
John.3SG bread.NA-OBV eat.VTA-THM.DIR-3SG.OBV

'John eats bread (animate).'

b. cân o-têm-a oskâtâskw-a mow-ê-yiwa
John.3SG 3.POSS-dog.NA-OBV carrot.NA-FUROBV eat.VTA-THM.DIR-3′.FUROBV

'John's (3SG) dog (OBV) eats a carrot (animate, FUROBV).'[6]

(8) VTI

a. cân wiyâs mîci-w
John.3SG meat.NI eat.VTI-3SG

'John eats meat (Inanimate).'

The {-w} in (8) is one of two third person suffixes in the VTIs, the other being {-Ø}. This morph is homophonous with third person markers in other conjugation classes. Alongside extensive person and direction morphology, several other categories may also

[6]As the marking for obviative and further obviative is formally the same, they must instead be distinguished on the basis of semantics and pragmatics.

be expressed on verbs.[7] Preverbs attach to the verb between person and the verb stem and serve several purposes. There are two types of preverbs: grammatical and lexical. The outermost of grammatical preverbs include those such as {ê-} and other Conjunct preverbs including {ka-}/{ta-}[8], and {kâ-}. While most preverbs are relatively freely combineable, these three are mutually exclusive. These morphs serve as complementizers and may have further functions, such as marking future or relative clauses. Closer to the verbal stem, one can observe another type of grammatical preverb for tense and aspect: {kî-} for past, {wî-} for prospective future, and {ka-/ta}- for definite future. Closer still to the verb are lexical preverbs, e.g., {kakwê-} 'try (to)', {nihtâ-} 'be good at', {nitawi-} 'go and (do something)', {âpihtâ-} 'half (of)/halfway', {kihci-} 'large', etc. (Wolfart, 1973, 1996; Wolvengrey, 2001), though even these show a gradience in lexicality/grammaticality.

2.3 Nêhiyawêwin Order

Algonquian languages are noted for their unique system of what is called Order, most easily recognized through allomorphy instantiated on the person-marking affixes of verbs. According to Bloomfield (1946, 97):

> The forms of the verb fall into five orders. Each order consists of one or more modes, each with a full set of forms. The Independent order takes prefixes; its principal mode, however, the indicative, has zero instead of *we-* for the third person. The other orders take no prefixes. The imperative has forms for second person actor only, and only one mode. The prohibitive has two modes with the same restriction, but also a third mode, the potential, with a full set of forms. The conjunct and interrogative orders are used only in subordinate clauses and as participles. The languages differ widely in their stock of modal forms; all seem to have lost a few, and some languages have created new ones.

The Orders described by Bloomfield are mutually exclusive. One can not have the morphology for both the Independent and the Conjunct, for example. For this reason, it

[7]For a large (though not yet complete) overview of Nêhiyawêwin morphemes (including common preverbs) see Cook and Muehlbauer (2010).

[8]This is a single morpheme that contains two allomorphs that are used in free variation. In central and southern Alberta, {ka-} is the more common form

seems obvious to group the mutually exclusive orders as a cohesive unit. As alluded to by Bloomfield, some Algonquian languages have fewer than five orders. Nêhiyawêwin is one of these languages, usually regarded as having only three orders: the Imperative, the Independent and the Conjunct. Despite the centrality of Order to the use of verbs, descriptions of the system as a whole remain vague for Nêhiyawêwin. Sometimes, Order is treated as a semantic alternation: the Imperative Order marks the imperative mood, while the Independent and Conjunct do not correspond to any specific mood. However, there is no such distinction between the Independent and Conjunct Orders. Instead, these Orders are usually analyzed through their morphological difference.

I argue that Order can be analyzed as an alternation. I suggest that Order as currently described is essentially two overlapping linguistic systems: one of mood/aspect and one of morphology that corresponds to a type of alternation previously undescribed: a paradigmatic alternation. To support this proposal, I will detail the morphological, syntactic, and semantic/pragmatic ways in which Order is used and defined and the ways in which these definitions are inadequate.

2.3.1 Morphology

Speaking strictly in terms of structural/morphological phenomena, the different Orders of Nêhiyawêwin can be divided into three main types: the Independent, the Conjunct, and the Imperative. The Independent is comprised of those forms which mark for any person argument and take a person prefix ({ni-} for first person, {ki-} for second, and no prefix for third person) and a set of suffixes (Bloomfield, 1946; Wolfart, 1973). The Conjunct is comprised of forms that also mark for any argument and which take no person prefix and one of a number of conjunct suffixes. The Imperative, on the other hand, marks for only person arguments includng the second person, cannot be used without such an arguement, does not make use of person prefixes, and uses a unique set of suffixes as compared to the Independent or Conjunct. Treating these Orders as of the same type due

to their mutual exclusivity, as done by Bloomfield (1946), results in a system depicted in Figure 2.1.

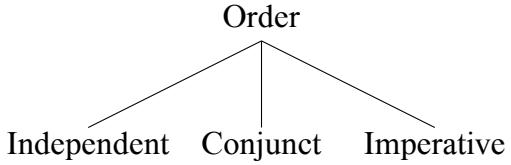

Figure 2.1: Order Ontology based on Morphology

This ontology, however, fails to capture a clear distinction of the Imperative from the Independent and the Conjunct. In a strictly structural sense, the shape of the Independent and Conjunct paradigms are similar to each other, while the Imperative's diverges from this standard substantially. To demonstrate and describe these differences, the structural makeup of the three canonical orders will be described below.

The Independent Order

According to Wolfart, the Independent order comes in two main forms: the preterit and non-preterit (1973). Preterit forms can be thought of as past-perfect constructions; conversely, the non-preterit form is essentially equivalent to the traditionally described indicative (Wolfart, 1973). Wolfart spends much of his description discussing the preterit forms of the Independent order, explaining the three types of preterit Independents. Since Wolfart's publication these preterit forms have largely fallen out of use in Nêhiyawêwin (Wolvengrey, 2011, 74) and so will not be further discussed. As previously mentioned, the Independent is identified by Bloomfield (1946), Wolfart (1973), and Cook (2014) as the Order that marks for all possible persons with the person prefixes {ni-} and {ki-} for first and second persons, respectively, and the lack of an prefix for the third and obviative persons. Independent forms are unable to take the {ê-} preverb (discussed later) which has begun to function primarily as a marker of Conjunct constructions.

Table 2.1 describes the Independent VII paradigm. Notice that only third person (and obviative) participants exist in this paradigm, and so no speech act participant prefix

Table 2.1: VII Independent Paradigm. Based on Wolvengrey (*2011, 413*).

	Prefix	Stem	Theme	SAP Person	Obviative	3SG	3PL	3’
3SG		mihkwâ				w		
3PL		mihkwâ				w	a	
3’SG		mihkwâ			yi	w		
3’PL		mihkwâ			yi	w	a	

Table 2.2: VTA Independent Direct, Mixed Participant Paradigm Excerpt. Based on Wolvengrey (*2011, 418*).

	Prefix	Stem	Theme	SAP Person	Obviative	3SG	3PL	3’
1	ni	wâpam	â			w	ak	
2SG	ki	wâpam	â			w	ak	
1PL	ni	wâpam	â	nân			ak	
21PL [9]	ki	wâpam	â	naw			ak	
2PL	ki	wâpam	â	wâw			ak	
3SG		wâpam	ê			w		
3PL		wâpam	ê			w	ak	
3’		wâpam	ê		yi	w		a

(SAP Person) or suffixes are used. These along with the final column, the additional third person obviative suffix, are unused but included to maintain consistency with the VAI, VTI, and VTA paradigms.

The VTA paradigms are further split. Here, a distinction is made between the *local* and *mixed* subsets. A *local* VTA subparadigm is one where the actor and the goal are both speech act participants (first or second persons), while the *mixed* subparadigm contains interactions between speech act participants and third or obviative persons. This subparadigm also contains third persons acting on obviative persons. This is presented in this way for the sake of convenience. In reality, one could place these non-speech act participant forms in their own sub-paradigm.

As seen in Tables 2.3 and 2.4, the paradigms of the VAI and VTI are extremely

[9]This represents the first person inclusive actor. In Algonquian linguistics, this is often considered as a second person form due to its morphology and its marking with the second person {ki-} prefix in the Independent.

similar, differing in their inclusion of a theme sign.[10]

Table 2.3: VAI Independent Paradigm. Based on Wolvengrey (2011, 415).

	Prefix	Stem	Theme	SAP Person	Obviative	3SG	3PL	3’
1SG	ni	nipâ		n				
2SG	ki	nipâ		n				
1PL	ni	nipâ		nân				
21PL	ki	nipâ		(nâ)naw				
2PL	ki	nipâ		nâwâw				
3SG		nipâ				w		
3PL		nipâ				w	ak	
3’		nipâ			yi	w		a

Table 2.4: VTI Independent Paradigm. Based on Wolvengrey (2011, 417). Note the difference of theme sign for local and non-local participants.

	Prefix	Stem	Theme	SAP Person	Obviative	3SG	3PL	3’
1SG	ni	wâpaht	ê	n				
2SG	ki	wâpaht	ê	n				
1PL	ni	wâpaht	ê	nân				
21PL	ki	wâpaht	ê	naw				
2PL	ki	wâpaht	ê	wâw				
3SG		wâpaht	am			(w)		
3PL		wâpaht	am			w	ak	
3’		wâpaht	am		(i)yi	w		a

In fact, there are some VAIs, like *âsokâham*, ‘s/he swims across’ that follow the general VTI paradigm and take the {-am} theme sign; conversely, some VTIs like *kâtâw*, ‘S/he hides something,’ take VAI morphology and follow the VAI paradigm. This has lead to an alternative interpretation of verb conjugation proposed by Wolvengrey (2011). Here, there is a three-way distinction between verbs based solely on the number of Animate participants: *V0* containing any verb forms with no Animate participants (corresponding to VII), *V1* containing verbs with only one Animate participant (corresponding to VAI and VTI), and *V2* containing verbs with two Animate participants (corresponding to VTA).

[10]Theme is used in the sense of traditional grammars, such as Goodwin (2002), where the theme sign

Table 2.5: VTA Independent Direct, Local Paradigm Excerpt. Adapted from Wolvengrey (*2011, 418*).

	Prefix	Stem	Theme	1SG	1PL	2PL
2SG	ki	wâpam	i	n		
2SG/PL	ki	wâpam	i	n	ân	
2PL	ki	wâpam	i			nâwâw

Table 2.6: VTA Independent Inverse, Mixed Participant Paradigm Excerpt. Wolvengrey (*2011, 418*).

	Prefix	Stem	Theme	SAP Person	Obviative	3rd Person	3PL	3’
1	ni	wâpam	ik(w)			w	ak	
2	ki	wâpam	ik(w)			w	ak	
1PL	ni	wâpam	ikw	inân			ak	
21PL	ki	wâpam	ikw	inaw			ak	
2PL	ki	wâpam	ikw	iwâw			ak	
3		wâpam	ik(w)			w		
3PL		wâpam	ik(w)			w	ak	
3’		wâpam	ikw		yi	w		a

Tables 2.2 and 2.5 through 2.7 gives a subset of an Independent VTA paradigm,[11] exemplifying direct and inverse forms for different pairs of participants for the VTA *wâpamêw* ‘s/he (animate) sees someone (animate)’. The person prefixes, and often the suffixes, remain the same while the direction morphology changes (note that some dialects allow for third person inverse forms with {-ikow} endings instead of {-ik}[12]). While the VTA Independent forms are decomposible, the Conjunct forms are not always so predictable.

is used to associate a stem with a particular paradigmatic shape.

[11]There are 36 person combinations in each of the Independent and Conjunct Orders, so not all pairings are presented in this dissertation.

[12]Note that the {iko} morph derives from the {ikw} morpheme along with an epenthetic /i/, the combination of which produces /iko/

Table 2.7: VTA Independent Inverse, Local Paradigm Excerpt. Based on Wolvengrey (2011, 418).

	Prefix	Stem	Theme	1SG/PL	2PL
2SG	ki	wâpam	iti	n	
2SG/PL	ki	wâpam	iti	nân	
2PL	ki	wâpam	iti		nâwâw

The Conjunct Order

Wolfart (1973) described four modes of the Conjunct, based on the presence or absence of the verb-final suffix {-ih} and the presence or absence of 'initial change' (an Algonquian process where the first vowel in the verb stem (or sometimes verbal prefixes) is mutated—abbreviated IC) (Wolfart, 1973). According to Wolfart those Conjunct verbs with both {-ih} and Initial Change are iterative and are named by him as such. Those without Initial Change but with {-ih} impart conditionality and are what Wolfart terms the *subjunctive*. Verbs with Initial Change but without {-ih} are simply called *Changed* and are the most commonly used Conjunct form, though Wolfart notes that Initial Change is beginning to fall out of use, being replaced instead by the use of the {ê-} preverb (1973). This view is consistent with that of Wolvengrey's account of {ê-} being born out of a regularization of a particular type of change, /i/ > /ê/, where the changed vowel was extracted from the construction to be used as a preverb, the verb stem retaining its original form (e.g., *itwêt* > *êtwêt* > *ê-itwêt*) (A. Wolvengrey, Personal Communication). Finally, those verbs without Initial Change or {-ih} are referred to as *simple* (Wolfart, 1973). A summary of this four way distinction is found in Table 2.8. In more contemporary Nêhiyawêwin orthography, the -ih ending is realized simply as a suffixal {-i}.

Cook (2014) provides further detail on the morphosyntactic and semantic behaviour of the Conjunct order. Agreeing with Wolfart (1973), Cook explains the widespread use of the order through several modes of the Conjunct. Unlike Wolfart's tetrachotamy, Cook gives a pentachotomy (2014). Under Cook's system, the Conjunct is split into the

Table 2.8: Wolfart's Conjunct modes. Adapted from Wolfart (1973, 45)

		Initial Change	
		+	-
/ih/	+	Iterative ('whenever it is')	Subjunctive ('if it be')
	-	Changed ('it being')	Simple ('that it is')

Table 2.9: Cook's Conjunct modes. Adapted from Cook (2014, 125)

Submode	Subtype	Form	Gloss
Changed	Changed Conjunct$_1$	ê-apiyân	'I sleep'
	Changed Conjunct$_2$	kâ-apiyân	'When I sleep'
	Iterative	êpiyâni	'Whenever I sleep'
Unchanged	Simple	ka-apiyân	'for him to eat'
	Subjunctive	apiyâni	'whenever I eat'

Changed and *Unchanged* modes (2014). The Changed Conjunct is further split into three subtypes: the *Changed Conjunct*$_1$, the *Changed Conjunct*$_2$, and the *Iterative Changed Conjunct*[13]. Although three subtypes are titled *Changed* due to being historically derived from changed forms, only the Iterative currently exhibits Initial Change. Changed$_1$ and Changed$_2$ on the other hand, are marked with the {ê-} and {kâ-} preverbs respectively [14]. The unchanged Conjunct forms are split into the *Subjunctive Simple Conjunct*, which are marked with no preverb and no Initial Change (but instead with a -i suffix appended to the person endings), and the *Irrealis Simple Conjunct*, which is marked with the {ka-} preverb. These forms are represented in Table 2.9.[15]

The following paradigms demonstrate the general shape of the Conjunct paradigm and represent the ê-Conjunct forms for the VII, VAI, VTI, and VTA conjunct classes.

[13]Where Wolfart (1973) identified an iterative/conditional morpheme as -ih, Cook (2014) follows the contemporary orthography.

[14]Wolfart (1973) classifies these two types together as changed conjunct forms, deriving {kâ-} from {kî-})

[15]Terminology for these terms vary between researchers. The subjunctive is sometimes referred to as the *future conditional* (Okimāsis, 2018; Ratt, 2016). Similarly, the term *timeless conditional* has been used in place of *iterative* (Harrigan et al., 2018).

As with the Independent paradigm, the VII Conjunct paradigm marks only for the third and obviative persons, as in Table 2.10.

Table 2.10: VII Conjunct Paradigm for mihkwâ, 'to be red'. Based on (*Wolvengrey, 2011, 413*)

	Prefix	Stem	Theme	SAP Person	Obviative	3rd Person	3PL	3'
3SG	ê-	mihkwâ				k		
3PL	ê-	mihkwâ				k	i	
3'SG	ê-	mihkwâ			yi	k		
3'PL	ê-	mihkwâ			yi	k	i	

Table 2.11: VAI Conjunct Paradigm for nipâ, 'to sleep'. Based on (*Wolvengrey, 2011, 415*).

	Prefix	Stem	Theme	SAP Person	Obviative	3rd Person	3PL	3'
1SG	ê	nipâ		yân				
2SG	ê	nipâ		yan				
1PL	ê	nipâ		yâhk				
21PL	ê	nipâ		yahk				
2PL	ê	nipâ		yêk				
3SG	ê	nipâ				t		
3PL	ê	nipâ				c	ik	
3'	ê	nipâ			yi	t		

Similar to the Independent, the Conjunct's VAI and VTI paradigms are strikingly similar. The main difference is the inclusion of an epenthetic /y/ in the SAP Person endings for the VAI paradigm, as well as the {-am} theme element in the VTI. These differences are exemplified in the differences between Tables 2.11 and 2.12.

The paradigmatic breakdowns used in Tables 2.13 through 2.16 highlight the theme morphs for the direct and inverse. There are alternative ways to analyze the endings in VTA paradigms, perhaps more straightforwardly by chunking all the suffixes together as sorts of portmanteau morphemes, as in Harrigan et al. (2017). For consistency and compatibility with Wolvengrey (2011), this dissertation will continue to use the paradigmatic patterns as presented in the four-conjugation class appendices of Wolvengrey (2011).

Table 2.12: VTI Independent Paradigm for wâpaht, 'to see it'. Based on (*Wolvengrey, 2011, 417*).

	Prefix	Stem	Theme	SAP Person	Obviative	3rd Person	3PL	3'
1SG	ê	wâpaht	am	ân				
2SG	ê	wâpaht	am	an				
1PL	ê	wâpaht	am	âhk				
21PL	ê	wâpaht	am	ahk				
2PL	ê	wâpaht	am	êk				
3SG	ê	wâpaht	am			k		
3PL	ê	wâpaht	am			k	ik	
3'	ê	wâpaht	am		(i)yi	t		

Table 2.13: VTA Conjunct Direct, Local Paradigm Excerpt for mow, 'to eat'. Based on Wolvengrey (*2011, 419*).

Actor → Goal	Prefix	Verb Stem	Theme	2SG/2PL	1PL
2SG → 1SG	ê-	mow	i	yan	
2SG/PL → 1PL	ê-	mow	i		yâhk
2PL → 1SG	ê-	mow	i	yêk	

Table 2.14: VTA Conjunct Inverse, Mixed Participant Paradigm Excerpt. Based on Wolvengrey (*2011, 419*).

Actor → Goal	Prefix	Verb Stem	Theme	Obviative	SAP	3SG	3PL
1SG → 3SG	ê-	mow			it		ik
2SG → 3SG	ê-	mow			isk		ik
3SG → 3'	ê-	mow	iko		yâhk		ik
1PL → 3SG	ê-	mow	iko		yahkw		ik
21PL → 3SG	ê-	mow	iko		yêkw		ik
2PL → 3SG	ê-	mow	iko			t	
3PL → 3'	ê-	mow	iko			t	ik
3' → 3"	ê-	mow	iko	yi		t	

Table 2.15: VTA Conjunct Inverse, Local Paradigm Excerpt. Based on Wolvengrey (*2011, 419*).

Actor → Goal	Prefix	Verb Stem	Theme	2SG/2PL	1PL
2SG → 1SG	ê-	mow	i	yan	
2SG/PL → 1PL	ê-	mow	i		yâhk
2PL → 1SG	ê-	mow	i	yêk	

Table 2.16: VTA Conjunct Direct, Mixed Participant Paradigm Excerpt. Based on Wolvengrey (2011, 419).

Actor → Goal	Prefix	Verb Stem	Theme	Obviative	SAP	3SG	3PL
1SG → 3SG	ê-	mow			ak		ik
2SG → 3SG	ê-	mow			at		ik
3SG → 3’	ê-	mow	â		yâhk		ik
1PL → 3SG	ê-	mow	â		yahkw		ik
21PL → 3SG	ê-	mow	â		yêkw		ik
2PL → 3SG	ê-	mow	â			t	
3PL → 3’	ê-	mow	â			t	ik
3’ → 3”	ê-	mow	â	yi		t	

The Imperative Order

Just as Bloomfield (1946) does, Wolfart (1973) describes two main Imperative modes: the Immediate and Delayed imperatives. The Immediate Imperative refers to a command to do something immediately, while the Delayed Imperative refers to a command to do something later. Because the Imperative only encodes command forms, both the immediate and the delayed mark only for second person forms. Resultingly, VII conjugation class of verbs, which only encodes third person and obviative actors, does not occur in the Imperative.

Across the remaining three conjugation classes, the Immediate Imperative describes an immediate command and is marked with no suffix, a {-tân} suffix, and a {-k} suffix for second person singular, first person inclusive, and second person plural, respectively. Again, the main differentiation between the VAI and VTI imperative paradigms is the latter containing a theme morph, as seen in Tables 2.17 and 2.18.

Additionally, the second person plural and all delayed forms contain an empenthetic /ɪ/. In each of these cases, the theme sign is realized as {-amw-} and the resulting /wɪ/ sequence coalesces to /o/, as in *wâpahtamok*, ‘See it, y’all!’. Where the {-amw-} and epenthetic /ɪ/ occur before an /h/, the surfacing form contains a long /o/, as in *wâpahtamôhkan*, ‘see it later!’

Table 2.17: VAI Imperative Paradigm. (*Wolvengrey, 2011, 395*)

	Verb Stem	Immediate	Delayed
2SG	nipâ		
21PL	nipâ	tân	
2PL	nipâ	k	
2SG	nipâ		hkan
21PL	nipâ		hkahk
2PL	nipâ		hkêk

Table 2.18: VTI Imperative Paradigm. (*Wolvengrey, 2011, 398*)

	Verb Stem	theme	Immediate	Delayed
2SG	wâpaht	a		
21PL	wâpaht	ê	tân	
2PL	wâpaht	amw	ik	
2SG	wâpaht	amw		ihkan
21PL	wâpaht	amw		ihkahk
2PL	wâpaht	amw		ihkêk

The Imperative paradigms for the VTAs looks somewhat different than the VAI and VTI paradigms. Because the VTAs take two Animate participants, the Imperative paradigm includes both first person and third person goals, as seen in Tables 2.19 and 2.20.

All forms except 2SG and 2PL acting on third persons in the Immediate imperative have a theme morph, {-â-} for the Mixed Participant Paradigm and {-i-} for the local. As in other cases, where one morpheme ends with /w/ and another begins with /i/, the surface form is realized as /o/, as in the Immediate second person acting on third singular for *wâpamihkok*, ‘witness them, y’all!’

Table 2.19: VTA Imperative Mixed Participant Paradigm (*Wolvengrey, 2011, 403*).

			Immediate		Delayed	
	Stem	Theme	3SG	3PL	3SG	3PL
2SG	mow		(i)	ik		
21PL	mow	â	tân	ik		
2PL	mow		ihkw	ik		
2SG	mow	â			hkan	ik
21PL	mow	â			hkahkw	ik
2PL	mow	â			hkêkw	ik

Table 2.20: VTA Imperative Local (*Wolvengrey, 2011, 403*).

			Immediate		Delayed	
	Stem	Theme	1SG	1PL	1SG	1PL
2SG	mow	i	n			
2SG/PL	mow	i		nân		
2PL	mow	i	k			
2SG	mow	i			hkan	
2SG/PL	mow	i				hkâhk
2PL	mow	i			hkêk	

Morphology Summarized

Morphologically, and in particular from a structural point of view, it is obvious that the Independent and the Conjunct have similar paradigmatic shapes. They each mark for the same persons and make use of similar prefixes (though the Conjunct does so more uniformly than the Independent) and suffixes to mark these persons. Conversely, the Imperative exhibits a far more restricted paradigm: among actors it marks only for second person and makes no use of person prefixes. Further, while the Independent and the Conjunct can occur in any verb class, the Imperative and VIIs are mutually exclusive. These factors, at least on their own, suggest an ontology that place the imperative separately from the Independent and Conjunct, which are more similar to each other. This is illustrated in Figure 2.2.

As will be seen throughout the rest of this chapter, this pattern of two orders being

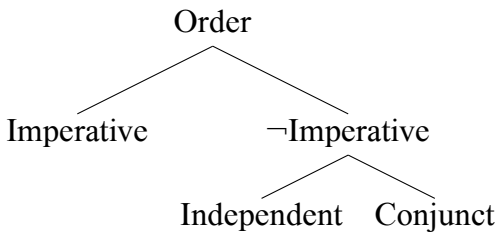

Figure 2.2: Morphological Ontology

similar while the remaining one stands apart is pervasive through various levels of representation. This poses difficulty for creating a description or analysis of Order as a unified tripartite system, as one order seems to act substantially different from the others.

2.3.2 Syntax

Progressing from Morphology, I will now discuss the syntax of the three canonical Nêhiyawêwin Orders. The syntactic differences between the Independent, Conjunct, and the Imperative orders are best described by Cook (2014). Although Wolfart (1973) touches on these differences, he does so without great detail. Wolfart (1973) mentions that while the Imperative and the Independent can stand alone (without a prior clause or referent), the Conjunct often represents some form of subordination (which requires another clause on which to depend). Further, he describes each of his four kinds of Conjunct forms as follows: the Simple Conjunct (without IC or a subjunctive suffix) generally follows future markers or conjunctions such as *nawac*, 'should', or *pitanê*, 'would that/may'; conversely, the Changed Conjunct (with IC but not a subjunctive suffix) indicates subordination with little other syntactic restrictions; the Iterative Conjunct (with both IC and the subjunctive suffix) generally occurs in narrative and participial clauses, and finally, the Subjunctive Conjunct (without IC but with a subjunctive suffix) represents some sort of conditionality and often futurity (Wolfart, 1973, 46). Similarly, Cook details the syntactic distribution of the Conjunct order, explaining like Wolvengrey (2011), that the Conjunct can occur in subordinate

(i.e. dependent clauses) (2014). In particular, Cook describes the Conjunct as *mostly* occurring in these subordinate clauses, but with her Changed Conjunct$_1$ class as additionally being possible in matrix clauses. A summary of Cook's Conjunct subtype distinction is found in Table 2.21 (2014, 125).

Table 2.21: Description of Conjunct Orders (*adapted from Cook* (*2014, 125*))

Submode	Subtype	Form	Matrix	Subordinate
Changed	Changed Conjunct$_1$	ê-apiyân	✓	✓
	Changed Conjunct$_2$	kâ-apiyân	✗	✓
	Iterative	êpiyâni	✗	✓
Unchanged	Simple	ka-apiyân	✗	✓
	Subjunctive	apiyâni	✗	✓

Although Cook explicitly does not discuss the Imperative, its syntactic distribution is similar to that of the Independent.

Cross linguistically, it has been reported that imperatives 'tend not to occur as dependent clauses' (Sadock and Zwicky, 1985, 174). Wolfart (1973) mentions that the imperative is often, but not exclusively, used along side a conditional clause, but in his examples, he gives only instances where the imperative verb is used in a matrix clause that contains a conditional subordinate clause. Alternatively, Lakoff (1984, 476) contends that Imperatives can occur in subordinate clauses provided the subordinate be introduced by *because* and the imperative actually convey a statement rather than an order. It is worth noting, however, that the evidence is provided for English, is not based in corpora or acceptability-judgement studies, and that the resulting 'grammatical' sentences (e.g. *I'm staying because consider the girl who pinched me*) are almost categorically ungrammatical to my ear. Takahashi (2008) presents a different approach, arguing that, at least in English, imperatives may be used as commands in certain concessive subordinate classes (e.g. *I am going to Toronto, although don't expect me to bring you anything back!*). Little has been written about this phenomenon in Nêhiyawêwin, and to do so would be beyond the scope of this dissertation. What can be said is that the Imperative is

not *exclusively* used in embedded clauses. This results in two organizational structures. The first patterns the Imperative syntactically with the Independent and the Changed Conjunct_1 as all three can occur in matrix clauses (note that the Changed Conjunct_1 also patterns with the other forms, as it can also be embedded), as in Figure 2.3.

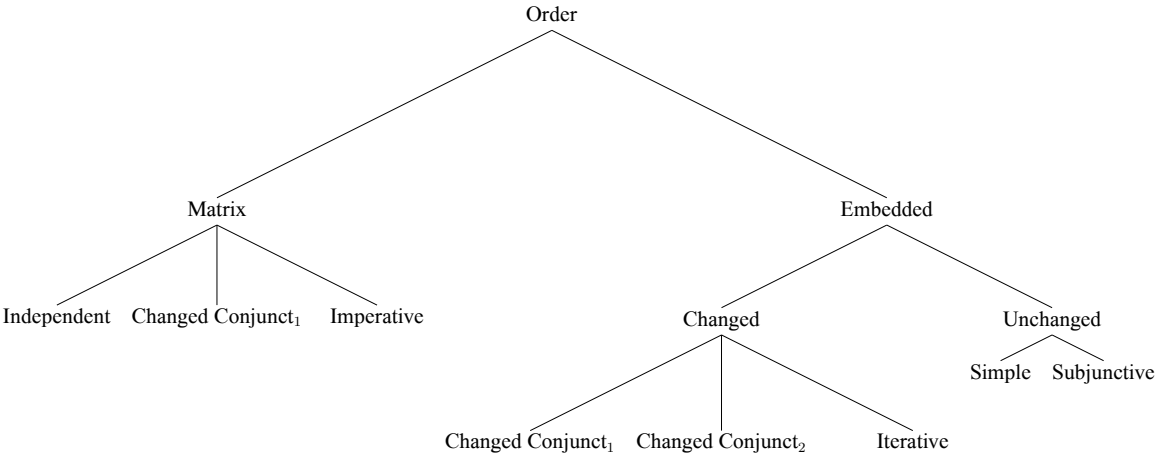

Figure 2.3: Syntactic Ontology 1

The second possibility is one where the Imperative occurs in both Matrix and Embedded clauses, as in Figure 2.4. In either of these situations, the syntactic system does not cleanly align with the morphological system of Order.

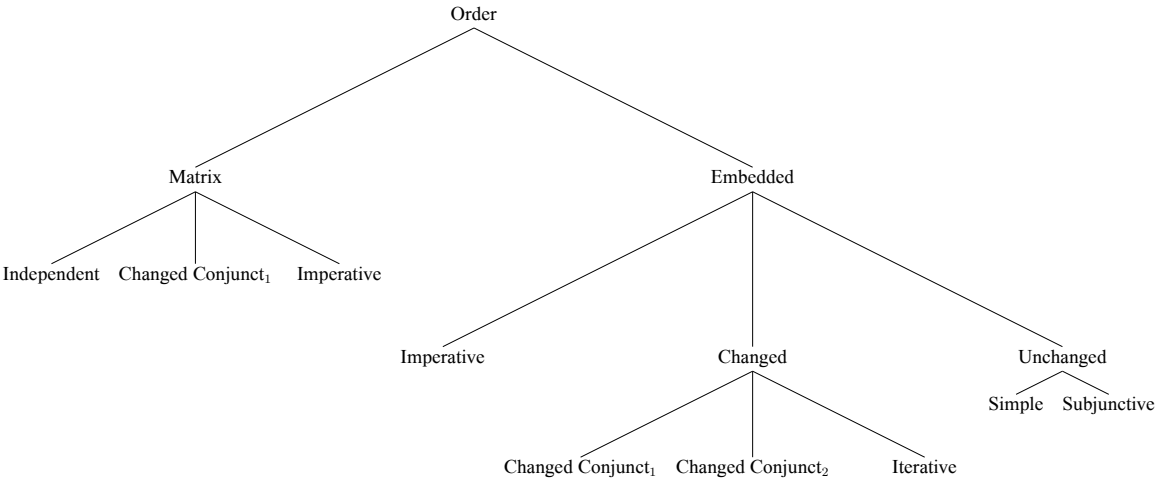

Figure 2.4: Syntactic Ontology 2

2.3.3 Semantics and Pragmatics

The semantics and pragmatics of Nêhiyawêwin Order can be broken down into two main theoretical constructs: (1) *sentence typing*, and (2) *clause typing*. Here, *sentence typing* refers to the three 'basic sentence types' as described by König and Siemund (2007), who identify the *declarative*, the *imperative*, and the *interrogative* as widespread typological phenomenon. These three Sentence Types are also represented in Nêhiyawêwin. While the Imperative order obviously corresponds to the imperative sentence type, the Independent and the Conjunct do not each represent one of the remaining sentence types. Instead, both the Independent and the Conjunct are able to be used as declarative constructions (in an unmarked or elsewhere case) as well as interrogatives (by making use of the {cî} clitic for the Independents or the Conjunct Order for content questions). This produces an ontology similar to the morphological organization seen previously, demonstrated in Figure 2.5.

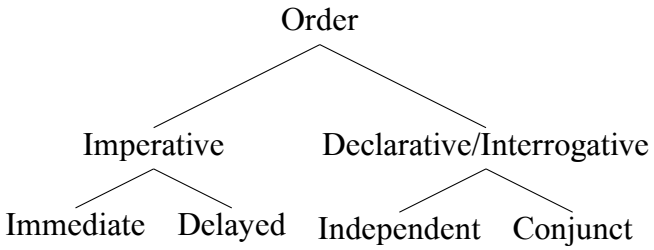

Figure 2.5: Semantic Ontology

For Cook (2014) the use of Order comes down to clause typing. Here, Cook (2014) distinguishes between indexical and anaphoric clauses. Indexical clauses are those that are grounded to the speech act, as the verb *kî-miyohtwâwak* is in (9). Indexical clauses are evaluated according the speaker as well as the time and place of the speech act; on the other hand, anaphoric clauses are evaluated according to some different anchor (Cook, 2014).

(9) mistahi **kî-miyohtwâ-wak** êkonik ôk âyisiyini-wak kâ-kî-ohpikih-iko-yâhkik
extremely PST-be.kind.VAI-3PL DIST.PL FOC.PL person-PL CNJ-PST-raise.VTA-INV.THM-3PL.1PL

'The people who raised us (...) **they were extremely good people.**' (Ahenakew, 2000, 38)

This is perhaps most clearly instantiated in the use of the {kî-} morph, which is used with past events. According to (Cook, 2014, 125), this past morph is interpreted in an unspecified way in Conjunct clauses, which Cook identifies as inherently anaphoric, but is interpreted with a strictly modal (and non-tense) meaning in the Independent. Cook (2014) describes these anaphoric clauses as being licensed by some antecedent, present in the discourse or in the real world knowledge of the interlocutors. Essentially, Cook (2014) describes anaphoric clauses as having *some* sort of semantic or syntactic relation with a licensor in another clause (as in (10)). She also contends that, in Nêhiyawêwin, anaphoric clauses are an elsewhere case that are defaulted to when an indexical clause is not present. The non-iteratve subjunctive form is not included by Cook, and its placement remains unclear.

(10) mistahi kî-miyohtwâ-wak êkonik ôk âyisiyini-wak **kâ-kî-ohpikih-iko-yâhkik**
extremely PST-be.kind.VAI-3PL DEM.PL FOC.PL person-PL CNJ-PST-raise.VTA-INV.THM-3PL.1PL

'**The people who raised us** (...) they were extremely good people.' (Ahenakew, 2000, 38)

Focusing specifically on the Conjunct modes, (Cook, 2014) distinguishes these forms by the ways in which their pragmatic/semantic propositions are introduced: the Changed Conjunct_2 and Iterative presuppose propositions, while Changed Conjunct_1 do not. Like the Changed Conjunct_1 forms, simple Conjuncts were not presuppositions, but are distinguished from Changed Conjunct_1 forms in that the latter are veridical statements, while simple Conjuncts are averidical Cook (2014, 302).[16] An adaptation of Cook's Order ontology is found in Figure 2.6.

Cook does not include the Imperative Order in her study, and it is difficult to determine where it would be placed in her ontology. Broadly, the imperative is clearly a clause type of its own: it represents an imperative clause as distinguished from declarative and interrogatives. If an indexical clauses is one that is rooted in the speech act, then the definition of *indexical* provided could just as easily apply to the Imperative Order. Indeed,

[16]It is unclear where Cook would place her subjunctive Conjunct in terms of veridicality, though given her placement of it as a type of 'simple conjunct', it seems possible that it would be an averidical form

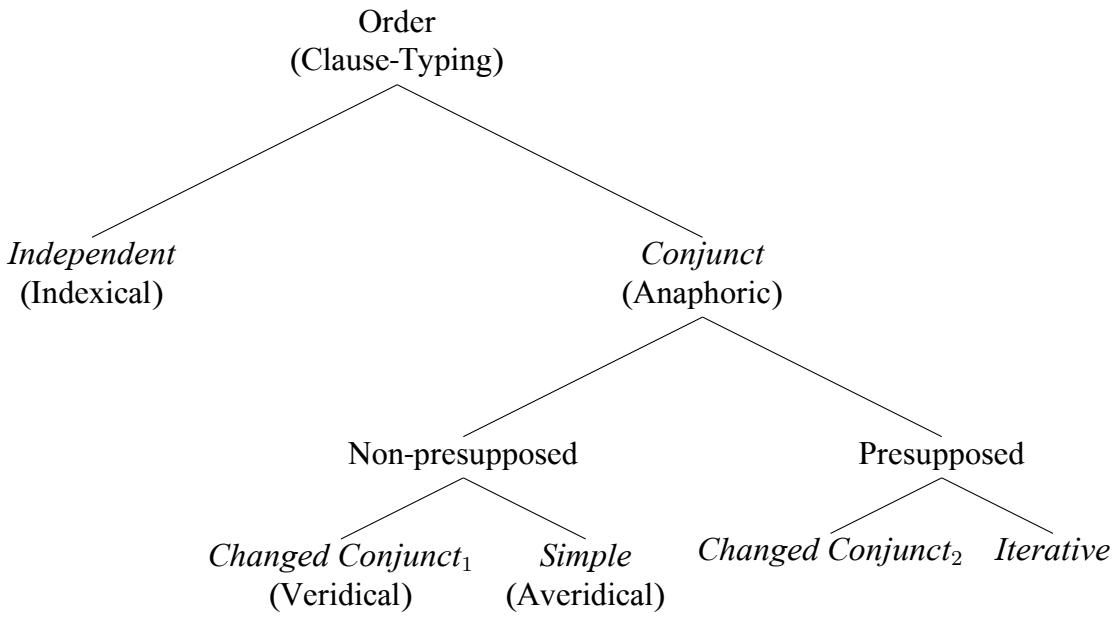

Figure 2.6: Order Ontology in Cook (*2014*)

Alcázar and Saltarelli (2014, 111) describes the Imperative (independent of any specific language) as "encoding the (indexical) parameters of the speech act, such as participant roles, temporality and locality". Under this analysis, we arrive at the ontology found in Figure 2.7.

Regardless of these interpretations, this sort of classification of Order treats the Independent, Conjunct, and Imperative not of the same kind (as is done in traditional descriptions of Algonquian grammar), but positions Conjunct as opposed to an Independent-Imperative conglomerate (distinct from descriptions by Bloomfield (1930), Wolfart (1973), Wolvengrey (2011) and others, which group the Independent and Conjunct together as opposed to the Imperative).

Conjunct modes in This Dissertation

As shown, while both agree that modes of the Conjunct exist, Wolfart (1973) and Cook (2014) vary in their descriptions of them. In order to study Order, it is critical to operationalize what different modes exist. Rather than simply taking either the Wolfart (1973) or Cook (2014), I opt to use corpus evidence to define the Conjunct modes on a

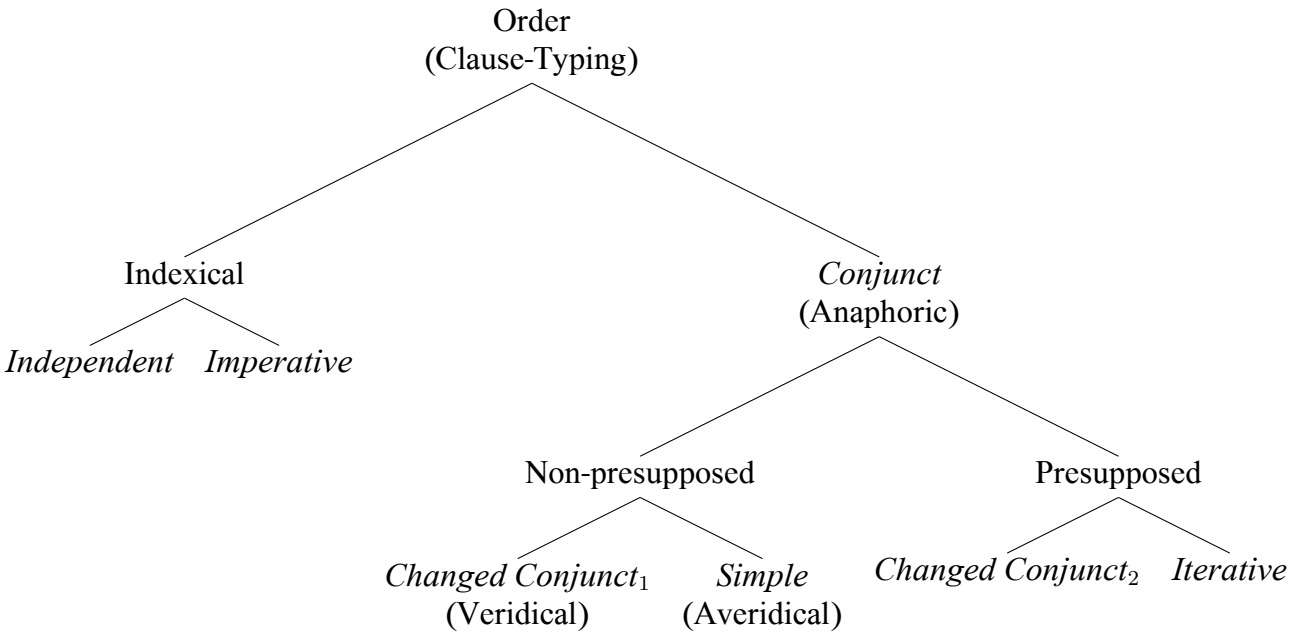

Figure 2.7: Order Ontology in Cook (*2014*)

structural basis. In the subset of the Ahenakew-Wolfart corpus (Arppe et al., 2020) used for this dissertation (see Chapter 4 for more detail), the following set of morphological patterns were found:

- ê- Initial (6373 tokens)
- ka-/ta- Initial (910 tokens)
- kâ- Initial (2458 tokens)
- Initial Change (54 tokens)
- Subjunctive {-i} (172 tokens)

Interestingly, there were no forms in the analyzed corpus that contained both a subjunctive suffix *and* IC (the *iterative* in Wolfart (1973) and Cook (2014)). While the corpus lacked an iterative, it did contain verbs with *only* IC,[17] a form seemingly missing in

[17]This may be due, at least regarding IC, to the fact that {ê-} was historically nothing more than a vehicle to indicate Initial Change (Wolfart, 1973, 46). In this way, one could consider the {ê-} prefixed Conjuncts as inherently Changed, though synchronically this is non-obvious. As a result, the remainder of this dissertation will not consider the {ê-} prefixed Conjuncts as examples of Initial Change.

Cook (2014). Further, the naming conventions used by Cook (2014) and Wolfart (1973) will not be used for this dissertation. Instead, I will refer to the Conjunct modes by their prefixes. The only exceptions to this are those forms where there is only initial change and those forms suffixed with the subjunctive morph. Because they can not be identified by a single prefix, they will be called the *Initial Change Conjunct* and the *Subjunctive Conjunct*.

In considering types of Conjunct, there is a structural difference between those types that have a grammatical, Conjunct specific, preverb such as {ê-}, {ka-}/{ta-}, and {kâ-}. These forms can be thought of as being *prefixed*, while the Initial Change and Subjunctive forms can be considered *bare*, due to their lack of a Conjunct prefix. Both Initial Change and Subjunctive forms have only a small number of tokens. Bare tokens with Conjunct endings but lacking either the Subjunctive {-i} or IC were excluded as contemporary speakers considered them as 'incorrect,' and their frequency in the corpus was even smaller than that of the Initial Change Conjunct.

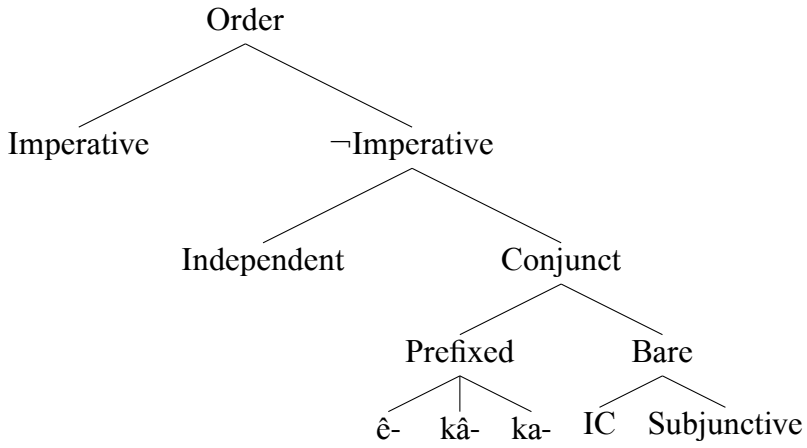

Figure 2.8: Morphological Ontology of Order

The most salient similarity between the Imperative, the Independent, and the Conjunct orders is that all three are able to inflect for second person items, at least in the non-VII classes. Beyond this, there are few similarities. Indeed, the Imperative differs from the other two Orders in that it:

- cannot be used with first, third, or obviative actors
- cannot take person-marking preverbs
- does not concern syntactic clause typing (and instead concerns speech-act level information)
- does not occur in statements of conditionality

Comparatively, the primary difference between the Independent and Conjucnt orders are the morphological exponents used in each Order. This results in a morphological system as visualised in Figure 2.8. Although the corpus used in this dissertation does not include Iterative Conjuncts, one could include them as a type of Subjunctive Conjunct (as both contain the Subjunctive suffix), resulting in the structure of Figure 2.9.

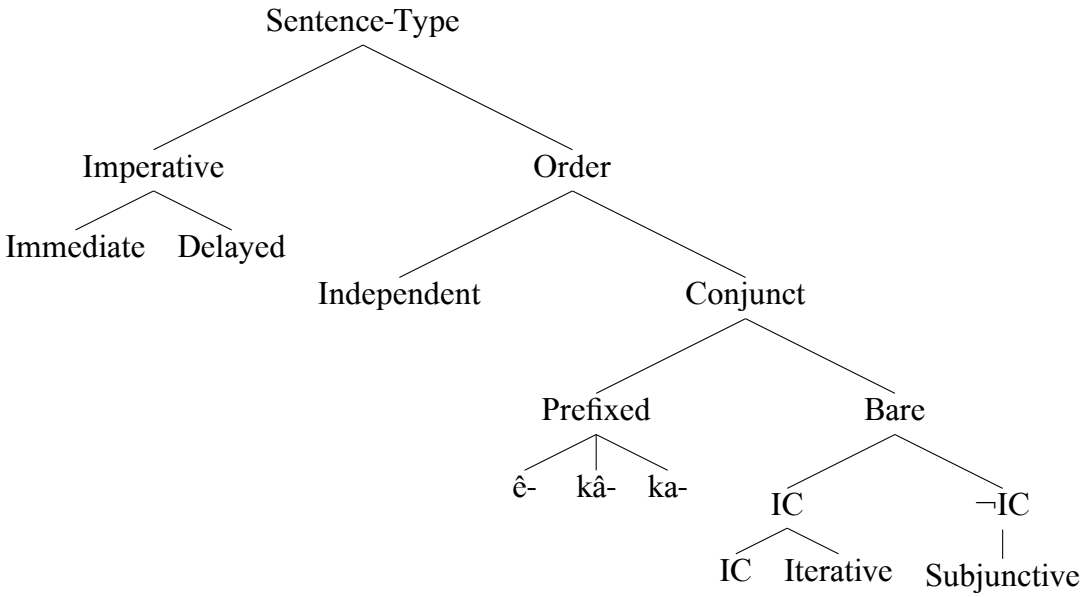

Figure 2.9: Morphological Ontology of Order 2

Alternatively, one could group the iterative with the Subjunctive, thus creating a Bare distinction between items with an {-i} suffix and those without, as in Figure 2.10.

There is no good theoretical reason to chose one of these options over the other. One could also choose to treat the Subjunctive, IC, and Iterative conjuncts as three separate

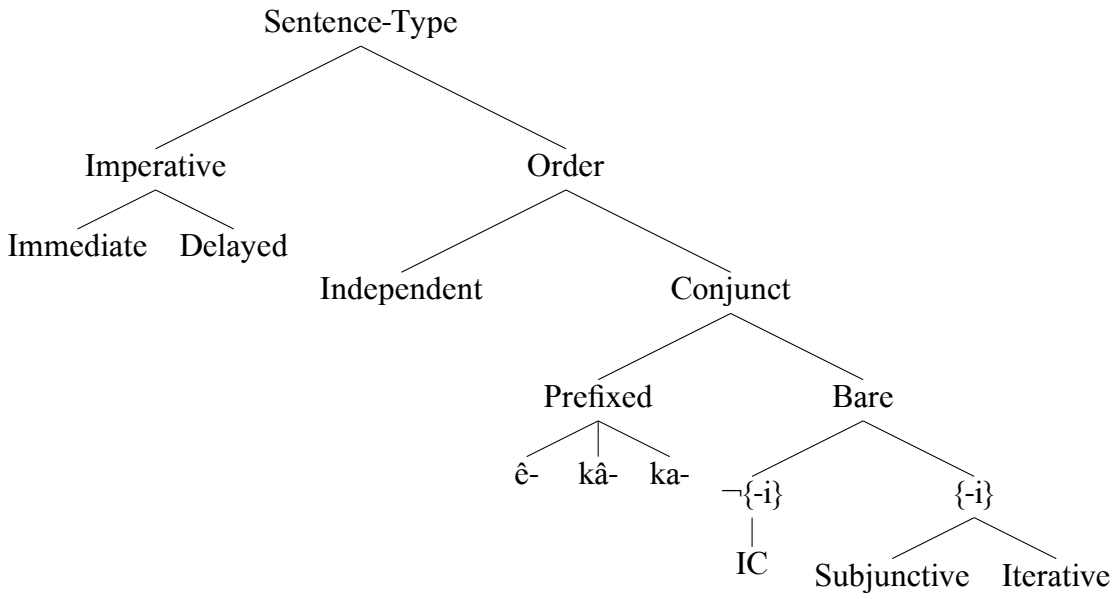

Figure 2.10: Morphological Ontology of Order 3

nodes, grouping none together. While this seems as valid as the previous two ontologies, it ignores the similarities of the Independent and the Conjunct in Nêhiyawêwin. In fact, because all bare forms in the language are combined for the sake of analysis in this dissertation, this distinction is not material for this dissertation.

2.3.4 Summary of Order

Nêhiyawêwin Order has been described as a system of linguistic features cross cutting various levels of representation. Morphologically, Order is a structural phenomenon which Algonquian languages use various exponents to mark person on verbs. Under this definition, we can identify three Orders:

1. Those where the VAI, VTI and VTA classes use circumfixes with {ni-} prefixes for first person and {ki-} prefixes for second person (the Independent)

2. Those with the prefixes {ê-}, {ka-}/{ta-}, {kâ-}, or Initial Change regardless of person (the Conjunct)

3. Those which use neither of these strategies (the Imperative)

This places the Independent and Conjunct together against the Imperative (which is essentially defined as not being Independent or Conjunct). Alternatively, we can identify two Orders:

1. Those that can mark for first, second, third, and obviative persons (the Independent and Conjunct)

2. Those that can mark only for the second person (the Imperative)

Again, in this situation the first of these proposed Orders would include what is traditionally called the Independent *and* what is traditionally called the Conjunct, with the second class making up the Imperative.

If we choose to define the phenomenon in terms of semantic, syntactic, and pragmatic behaviour, we can refer to Figure 2.7, wherein Independent and Imperative are indexical, while the Conjunct in anaphoric. Contrary to the previous descriptions, this places Conjunct apart from the other Orders.

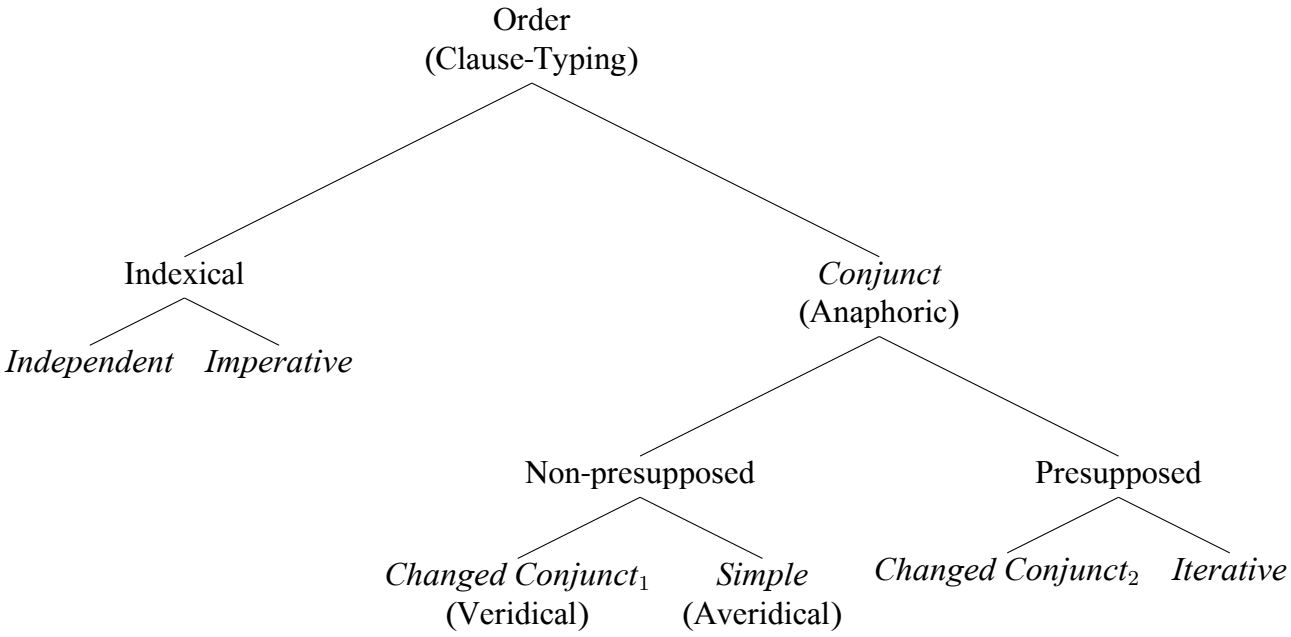

Figure 2.7: Order Ontology in Cook (*2014*) (*repeated from page 31*)

Finally, if we consider Order purely in terms of semantics, we can define Order as a

system of distinguishing mood (the imperative vs. the declarative). In this classification, the Independent and Conjunct are not distinguished by mood in the same way that they can be contrasted against the Imperative (cf. 2.11).

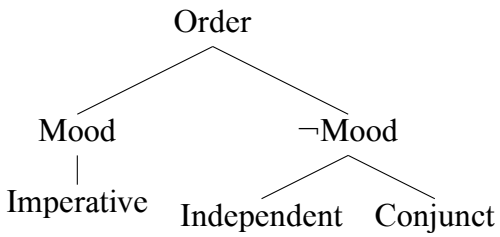

Figure 2.11: Semantic Order Ontology

Thus, we again have a situation where the Imperative is of a different kind than the Independent/Conjunct. Regardless of what scheme one uses to describe Order in Nêhiyawêwin, there is no way to divide the Independent, Imperative, and Conjunct such that they're all of the same kind or on the same level. The best argument for equating the Independent, Conjunct, and Imperative is that occurrence in one of these precludes occurrence in the other (i.e. there is no such thing as an Independent Imperative).[18] Under this definition, a tri-partite Order is essentially an operation that takes a verb stem, a linguistic person or persons, and a direction (if needed) and produces a surface form as in equation (2.1).[19]

$$\text{Conjunct}(\text{wâpam}, (1\text{SG}, 3\text{SG}), \text{inverse}) = \text{ê-wâpamit} \tag{2.1}$$

In this way, one can think of Order as an operation that applies to verbs; however, the Imperative is incompatible with the VII class, while both the Independent and Conjunct can apply to any class. Even considering Order as this sort of formal function leads to a distinction between the Imperative and the Independent/Conjunct. In terms of structure, behaviour, and semantics, this difference persists. This conflict is problematic to the study

[18]Unless, of course, one identifies the Delayed Imperative as a Conjunct form for the Imperative, as described above.

[19]I make no claim of psychological reality in this statement, it is purely metaphorical.

of Nêhiyawêwin grammar, as any claim about Order needs to be relevant to all three of these categories. For these reasons, the use of the term *Order* will be redefined in this dissertation. Instead of creating a three way split between Independent, Conjunct, and Imperative, I will consider Order to be a grouping of allomorphic alternations in the paradigm. Therefore, this dissertation will refer only to the Independent and Conjunct.

In terms of describing what the Imperative is, if not Order, I propose that the Imperative is a construction that acts as an illocutionary force indicating device (Searle and Vanderveken, 1985) marking a command. Under this system, we can understand the interrogative to be marked through the use of the {cî} morph, and the declarative to remain unmarked. Thus the concept of mood (which is mostly imparted by preverbs in Nêhiyawêwin) is made separate from the idea of Order entirely. Thus, while the Independent and Conjunct may still be referred to as Order, the Imperative is of the sentence-type or illocutionary force.

2.3.5 Alternation

By extricating the Imperative from the system of Order, we are left with a binary distinction of Independent and Conjunct. This juxtaposition presents two ways of encoding person number with different morphemes. In other words, while the shape and grammatical content (e.g. both Orders mark exactly the same persons) of the *paradigms* are the same, the actual exponents that are realized in the cells are not. According to Cook (2014), these two alternatives correspond one-to-one to clause typings, the indexical and the anaphoric. This view of Order as two alternative constructions used to encode different meaning is essentially one of *alternation*.

In its broadest conception, the idea of an alternation is simply one in which some linguistic form—be it phonological, morphological, syntactic, or other—is contrasted with another. Pijpops (2020) presents an overview of the concept covering the three traditional definitions (1-3) along with three more recently developed conceptions (4-6):

1. Alternations share meaning, are similarly processed in the mind, and vary dialectally.

2. Alternations share the same meaning, do **not** vary according to dialect, but **are** differently processed in the mind.

3. Alternations have a difference in meaning that varies due to some lexical influence

4. Alternations represent any point where the speaker must make a choice in what is said

5. Alternations are a tool to analyze phenomenon that a linguist deems interesting

6. Alternations are items with special theoretical relations to one another

In addition to Pijpops (2020)'s definitions, there are other ways to approach alternations. Specifically, one can make use of a lexicographically grounded approach which considers the concept of synonymy and the way in which synonyms and near synonyms can be used in similar (but not identical) contexts. In this vein, Cruse (2000, 156) discusses the concept of synonymy, which he defines as not simply words with the same meaning, but "words whose semantic similarities are more salient than their differences." In particular, Cruse identifies three types of synonyms: absolute synonyms (which are fully equivalent and occur rarely), propositional synonyms, (which alternate without changing the truth condition of a statement, but which may differ in speaker attitude or register), and plesionyms/near-synonyms (which can be said to share core semantic properties, even if they differ in 'minor' or 'background' ways) (Cruse, 2000, 157-159). Because any of these forms of synonymy necessarily concern the employment of one of many forms for the same referent, synonymy is a clear case of alternation. Similarly, Inkpen and Hirst (2006) describe near-synonymy as words which can not be chosen between without knowledge of contextual differences. Following from this

lexicographic approach, alternations can be construed on various levels: conceptual-semantic alternation, stylistic-semantic alternation, and a syntactic-semantic alternation (Arppe 2008, 8; cf. Edmonds and Hirst 2002 for an earlier discussion of a similar concept). According to Arppe (2008, 8), conceptual-semantic alternations concern words that mean generally the same thing and can be used (roughly) interchangeably (e.g. *dash* and *sprint*); stylistic-semantic alternations occur between words or phrases that share similar meanings, but contain different connotations (*poop* and *shit*); and syntactic-semantic alternations deal with similar utterances which take different syntactic patterns (*comb* (*through*) and *inspect*). These levels of representation consider alternations as near-synonymous sets that can make use of three latter definitions presented by Pijpops (2020), particularly as a point-of-choice. They also roughly correspond to those of Hanks' lexical, semantic, and syntactic-type alternations (2013, 173). Arppe (2008, 10) also proposes a subset of syntactic-semantic alternations referred to as constructional alternations. These concern phrases instead of words, keep the same central meaning, though which may differ in more subtle, often pragmatic dimensions. This type of alternation is discussed in Biber et al. (1998) and mentioned as a caution for taxonomic classification by DiMarco et al. (1993). Framing a phenomenon as an alternation creates a structured difference that researchers can investigate. As an example, discussing the Independent and Conjunct Orders as an alternation allows researchers the ability to compare and contrast the particular morphological processes that go into each cell in the paradigm. This is possible precisely because the paradigms alternate straightforwardly while the general size and shape of the paradigm remain constant.

Using the lens of alternation, I propose that Order can be studied with systematic quantitative methods, as put forth by Arppe (2008) and Arppe (2009), expanding from Gries (2003) and Bresnan et al. (2007). In Arppe (2008), various Finnish synonyms for *think* in a corpus are analyzed for their morphological, syntactic, and semantic values. Each token is given a tag set that summarizes these features, and a multivariate statistical

analysis technique such as logistic regression is used to determine which features predict the use of which synonym (e.g. that the use of a *think* verb in a direct quote significantly increases the likelihood of the use of *miettiä*, 'think, ponder'). I suggest that Order could be studied in a similarly principled way: instead of considering the alternation between two synonymous lexemes, I consider near-synonymous inflections. Related work, such as that by Divjak and Gries (2006) who investigated nearly synonymous *try* verbs in Russian; Klavan and Divjak (2016) who reviewed nearly synonymous choice in Arabic, English, Estonian, and Russian through statistical methods; and Klavan (2020) who approached Estonian near-synonymy with both logistic regression and a related technique, naïve discriminitive learning, further motivate this research.

However, viewing Order as an alternation can be difficult given the above definitions. Order cannot be conceived as a conceptual-semantic alternation as the Independent and the Conjunct *do* appear to have some restrictions on their syntactic distribution; similarly, Order cannot be said to be a stylistic-semantic alternation as there is no such connotation difference; Order also cannot be considered either syntactic-semantic or constructional alternations as these fail to capture that the alternation applies not to a set of lexemes or construction frames, but an entire morphological paradigm. Thus, I argue that the phenomenon of binary Order, between Independent and Conjunct, is a form of nearly synonymous constructional alternation, but one that has remained, as of yet, undescribed. I propose that order represents a *paradigmatic alternation*. A paradigmatic alternation is here defined as one where *any* lexeme of a particular word class is able to take two or more different paradigms but where each of those paradigms is identical in shape but different in exponence. This differs from similar phenomena such as noun class where there are indeed alternating paradigms with similar or identical shapes differing in exponents but where it is not the case that any noun can occur in any paradigm. Instead, the paradigm which a lexeme occurs in is functionally an attribute of a lexeme. [20]

[20]It is worth pointing out that this type of alternation is not necessarily unique to Nêhiyawêwin. In fact, a very similar pattern of inflection is seen in Koiari tense. This is further discussed in Harrigan and Arppe

Viewing Order as an alternation allows for the investigation of how Orders behave. Rather than relying on impressionistic analyses of these phenomena, researchers can construe this alternation as a question of binary classification. Building off similar work by (Arppe, 2008), one can focus on Order identity itself as a response variable that is predicted by a number of morphosyntactic and semantic features.

Viewed as an analysis of an alternation, the primary research question of this dissertation is as follows: what morphosyntactic and semantic features affect a lemma's propensity to occur in a particular alternation of Order or mode? Adopting a usage-based approach based in the distributional hypothesis (Firth, 1957; Harris, 1954), this research will utilize quantitative methodologies in an effort to see to what extent empirical, corpus-based evidence can guide us in understanding Nêhiyawêwin Order.

For the purposes of this dissertation, three main levels of paradigmatic alternation will be considered. The first of these alternations is the **Independent vs. Conjunct** alternation. This is the highest level alternation and is essentially that of the phenomenon of Order. This alternation thus represents the high level decision of what morphological paradigm is to be used. The second alternation is the **Independent vs. ê-Conjunct**. Although this alternation appears to cross multiple levels of representation (e.g. the decision to use an Independent vs. a Conjunct form appears to precede the decision to use an ê-Conjunct), the linguistic motivation for this alternation is found in the similar behaviour and functions as described by Wolfart (1973) and Cook (2014). The final alternation is the **Conjunct Type** alternation between the ê-Conjuncts, kâ-Conjuncts, and all other Conjunct types. This alternation is perhaps the most straightforward, and is motivated by the fact that one must choose what form of Conjunct they use for a verb.

(Forthcoming)

Chapter 3

Automatic Semantic Classification

Abstract

Previous versions of this paper were published as Harrigan and Arppe (2021) and Harrigan and Arppe (2023). This paper details a semi-automatic method of word clustering for the Algonquian language, Nêhiyawêwin (Plains Cree). Although this method worked well, particularly for nouns, it required some amount of manual postprocessing. The main benefit of this approach over implementing an existing classification ontology is that this method approaches the language from an endogenous point of view, while performing classification quicker than in a fully manual context.

3.1 Introduction

Grouping words into semantic subclasses within a part of speech is a technique used widely throughout quantitative and predictive studies in the field of linguistics. Bresnan et al. (2007) use high level verb classes to predict the English dative alternation, Arppe (2008) uses verb class as one of the feature sets to help predict the alternation of Finnish *think* verbs, and Yu et al. (2017) use polarity classifications (*good* vs. *bad*) from pre-defined lexica such as WordNet Miller (1998). In many cases, classifications within word classes allow researchers to group words into smaller cohesive groups to allow for use as predictors in modelling. Rather than using thousands individual lexemes as predictors, one can use a word's class to generalize over the semantic features of individual lexemes to allow for significantly more statistical power.

While extensive ontologies of word classifications exist for majority languages like English (Miller, 1998), German (Hamp and Feldweg, 1997), and Chinese (Wang and Bond, 2013), minority languages, especially lesser resourced languages in North America generally do not boast such resources.[1] Where such ontologies do exist, for example in Innu-aimun, also known as Eastern Cree (Visitor et al., 2013), they are often manually created, an expensive process in terms of time. Alternatively, they may be based upon English ontologies such as WordNet. This opens the window to near-automatic ontology creation by associating definitions in a target language and English through a variety of methods. This is especially important, given the amount of time and effort that goes into manually classifying a lexicon through either an existing ontology (be it something like Rapidwords[2] or even Levin's like classes (Levin, 1993)). Moreover, there is a motivation based in understanding a language and its lexicalization process on its own terms, though how to do this with a lesser resourced language remains unclear.

[1]There is one attempt at semantically classifying Nêhiyawêwin through automatic means found in Dacanay et al. (2021). This work makes use of similar techniques as described in this paper, differing mainly in its mapping of Nêhiyawêwin words onto Wordnet classes.

[2]See http://rapidwords.net/

3.2 Background

I began word classification in preparation for modelling a morpho-syntactic alternation in Nêhiyawêwin verbs. One hypothesis I developed for this alternation, based on Arppe (2008), is that the semantic classes of the verbs themselves as well as their nominal arguments would inform the verbal alternation. Due to constraints of time, I investigated methods to automatically classify both verbs and nouns in Nêhiyawêwin. Although statistical modelling remains the immediate motivator for the authors, semantic/thematic classifications have a wide range of benefits for language learners and revitalization, particularly in online lexicographic resources, where one may want to view all words to do with a theme, rather than simply finding translations of single English words.

In creating a framework for automatic semantic classification, I make use of Word2vec (Mikolov et al., 2013a) word embeddings. Word embeddings are words represented by *n*-dimensional vectors. These vectors are ultimately derived from a word's context in some corpus through the Word2vec algorithm. Unfortunately, the Word2vec method is sensitive to corpus size. I initially attempted to create basic word and feature co-occurrence matrices based on a 140,000 token Nêhiyawêwin corpus Arppe et al. (2020) to create word vectors using Principal Components Analysis, but in the end found the results to be not practically useful. Similarly, an attempt at both tf-idf and Word2Vec using only the Nêhiyawêwin dictionary produces mostly ill-formed groupings, though in these cases preprocessing by splitting verbs and nouns was not performed. Regardless, the poor performance was most certainly due simply to the paucity of data. Although the available corpora are small, Nêhiyawêwin does have several English-to-Nêhiyawêwin dictionaries, the largest being Wolvengrey (2001). Although a bilingual Nêhiyawêwin-English dictionary, it is one formed from an Indigenous point of view, based on vocabulary from previous Nêhiyawêwin language resources, some of which have been compiled by Nêhiyawêwin communities from their own perspectives, or gleaned from a number of texts collections rather than attempting to find Nêhiyawêwin word matches for

a pre-defined set of English words. This results in dictionary entries such as sakapwêw: it roasts over a fire (by hanging, with string on stick). Definitions such as this take into account the nuanced cultural understanding reflected in the word's morphology.

3.3 Methodology

To address the issue of corpus size, I attempted to bootstrap my classification scheme with pre-trained English vectors in the form of the 3 million word Google News Corpus, which represents every word with a 300-dimensional vector.[3] I make use of the English definitions (sometimes also referred to as glosses) provided in Wolvengrey (2001) and fit to each word its respective Google News Corpus vector. This dictionary makes use of lemmas as headwords, and contains, at the time of writing, 21,717 entries. The presumption is that the real-world referents (at least in terms of denotation) of English and Nêhiyawêwin words are approximately comparable, in particular when taking the entire set of words in an English definition. Stop words (common words that supply little lexical or semantic information) were removed, and where content words were present in definitions in Wolvengrey (2001) but *not* available in the Google News Corpus, synonyms were used (one such example might be the word *mitêwin*, which is unavailable in the corpus and thus would be replaced with something like *medicine lodge* or deleted if a synonym was given in the definition as well). Because the Google News Corpus is based in American spelling, while Wolvengrey (2001) is based in Canadian spelling, American forms (e.g. *color, gray*) were converted into Canadian forms (e.g. *colour, grey*). If such preprocessing is not performed, these words are simply unavailable for clustering, as they lack a matching vector.[4] Where a Nêhiyawêwin word had more than

[3]This corpus was trained on a large corpus of 100 billion words. Available at https://code.google.com/archive/p/word2vec/

[4]In reality, there were only a handful of cases where words occurred in the dictionary but not in the Google News Corpus. Because there are so few examples of this, even simply leaving these items out would not substantiqally change clustering results.

one word sense, each sense was given a separate entry and the second entry was marked with a unique identifier. Finally, where needed, words in the Nêhiyawêwin definitions were lemmatized.

Once every word in Wolvengrey (2001) definitions matched an entry in the Google News Corpus, I associated each word in a Nêhiyawêwin definition with its respective Google News Vector. That is, given a definition such as awâsisihkânis: small doll, the resulting structure would be:

$$\text{awâsisihkânis} = \begin{bmatrix} 0.159 \\ 0.096 \\ -0.125 \\ \vdots \end{bmatrix} \begin{bmatrix} 0.108 \\ 0.031 \\ -0.034 \\ \vdots \end{bmatrix}$$

Because all word-vectors in the Google News Corpus are of the same dimensionality, I then took the resulting definition and averaged, per dimension, the values of all its constituent word-vectors. This produced a single 300-dimensional vector that acts as a sort of naive sentence vector for each of the English glosses/definitions:

$$\text{awâsisihkânis} = \begin{bmatrix} 0.134 \\ 0.064 \\ -0.080 \\ \vdots \end{bmatrix}$$

Mikolov et al. (2013b) mention this sort of naive representation and suggests the use of phrase vectors instead of word vectors to address the representation of non-compositional idioms; however, given the way Wolvengrey (2001)'s definitions are written (e.g. with few idiomatic or metaphorical constructions), and for reasons of computational simplicity, I opted to use the above naive implementation in this paper.

After creating the sentence (or English definition) vectors, I proceeded to cluster

definitions with similar vectors together. To achieve this, I created a Euclidean distance matrix from the sentence vectors and made use of the hclust package in R (R Core Team, 2017) to perform hierarchical agglomerative clustering using the Ward method (based on the experience of Arppe (2008) in using the method to produce multiple levels of smaller, spherical clusters). This form of clustering is essentially a bottom-up approach where groupings are made by starting with individual labels with the shortest distance, then iteratively at a higher level making use of the clusters that result from the previous step or remaining individual levels; this second step is repeated until there is a single cluster containing all labels. This method of clustering creates a cluster tree that can be cut at any specified level after the analysis has been completed to select different numbers of clusters, allowing researchers some degree of flexibility without needing to rerun the clustering. This method is very similar to what has been done by both Arppe (2008), Bresnan et al. (2007), and Divjak and Gries (2006). The choice of how many clusters were used was based on an impressionistic overview of effectiveness by myself.

For my purpose, I focused on the semantic classification of Nêhiyawêwin nouns and verbs. Nêhiyawêwin verbs are naturally morphosemantically divided into four separate classes: Intransitive verbs with a single inanimate argument (VII), Intransitive verbs with a single animate argument (VAI), transitive verbs with an animate actor[5] and an inanimate goal (VTI), and verbs with an animate actor and goal (VTA). For verbs, clustering took place within each of these proto-classes. Among the VIIs, 10 classes proved optimal, VAIs had 25 classes, VTIs with 15 classes, and VTAs with 20 classes. The choice to preprocess verbs into these four classes was chosen as not doing so resulted in a clustering pattern that focused mainly on the difference between transitivity and the animacy of arguments. Any more or fewer classes and HAC clusters were far less cohesive with obvious semantic units being dispersed among many classes or split into multiple classes

[5]As discussed in Wolvengrey (2005), Nêhiyawêwin sentences are devoid subject and objects in the usual sense. Instead, syntactic roles are defined by verbal direction alignment. For this reason, I use the terms *actor* and *goal* instead of *subject* and *object*.

with no obvious differentiation. Similarly, verbs were split from nouns in this process because definitions in Wolvengrey (2001) vary significantly between verbs and nouns.

Nouns are naturally divided into two main classes in Nêhiyawêwin: animate and inanimate.[6] For my purpose I divide these further within each class between independent (i.e. alienable) and dependent (i.e. inalienable) nouns to create four main classes: Independent Animate Nouns (NA), Dependent Animate Nouns (NDA), Independent Inanimate Nouns (NI), and Dependent Inanimate Nouns (NDI). The reason for this further division is due to the morphosemantic differences between independent and dependent nouns in Nêhiyawêwin. While independent nouns can stand on their own and represent a variety of entities, dependent nounds are semantically and morphologically dependent on some possessor. I opted to pre-split NDIs and NDAs into their own classes, so as not to have the clustering focus on alienablity as the most major difference.[7]

3.4 Results

In all cases, clusters produced by this procedure needed some amount of post-processing. For nouns, this post-processing was minimal and mostly took the form of adjustments to the produced clusters: moving some items from one class to another, splitting a class that had clear semantic divisions, etc. For the verbs, this processing was often more complex, especially for the VAI and VTA classes. Although most clusters produced somewhat cohesive semantic units, the largest clusters for the VAI and VTA classes acted as, essentially, catch-all clusters. Although computationally they seemed to have similar vector semantics, the relationship between items was not obvious to the human eye. Postprocessing for these clusters took more time than other classes and essentially

[6]Although this gender dichotomy is *mostly* semantically motivated (e.g. nouns that are semantically inanimate are part of the inanimate gender) this is not always the case as in the word *pahkwêsikan*, 'bread', a grammatically animate word.

[7]Preliminary results for words not seperated by their conjugation class or declension did, in fact, create clusters based around these obvious differences. This likely due to the way definitions were phrased (e.g. dependent nouns would have a possessive determiner or pronoun).

Table 3.1: HAC built cluster counts vs. counts after postprocessing

	HAC classes	Manually Adjusted Classes	Lexemes
VII	10	6	581
VAI	25	13	5254
VTI	15	6	1825
VTA	20	7	1781
NI	15	13	3650
NDI	3	2	245
NA	10	8	1676
NDA	3	3	191

composed of using the more cohesive clusters as a scaffold into which one may fit words from these catch-all clusters. In most cases, this resulted in very slightly more clusters after postprocessing, though for VAIs this number was significantly higher, and for the NDIs it was slightly lower. Table 1 lists the number of clusters directly from HAC and from postprocessing. The actual quality of clustering varied from class to class. In general, nouns resulted in much more cohesive clusters out-of-the-box and required far less postprocessing. For example, nearly all NI_{14} items referred to parts of human bodies (and those that did not fit this description were terms clearly related to, or containing, body parts like aspatâskwahpisowin, 'back rest'), NI_{13} was made up of trapping/hunting words and words for nests/animals.

The NA classes produced through HAC were similarly straightforward: NA_9 was made up of words for trees, poles, sticks, and plants; NA_8 was made up entirely of words relating to beasts of burden, carts, wheels, etc.; while much of NA_3 and NA_7, and nearly all of NA_2 referred to other animals. Once manually postprocessed, the NA lexemes settled into 8 classes: $\mathrm{NA\text{-}persons}$, $\mathrm{NA\text{-}beast\text{-}of\text{-}burden}$, $\mathrm{NA\text{-}food}$, $\mathrm{NA\text{-}celestial}$, $\mathrm{NA\text{-}body\text{-}part}$, $\mathrm{NA\text{-}religion}$, $\mathrm{NA\text{-}money/count}$, and $\mathrm{NA\text{-}shield}$.[8]

The NDI and NDA classes required almost no postprocessing: NDA_1 and NDA_3 were each made up of various family and non-family-based relationships, while NDA_2 was

[8]This class refers to forms such as *nakahâskwân* and *pahpahâhkwân*, which both translate as 'shield', despite being grammatically animate.

made up of words for body parts and clothing. The resulting classes for these were: NDA-Relations, NDA-Body, and NDA-Clothing.

The NDI lexemes took two classes: the vast majority of NDI forms referred to bodies and body parts while two lexemes referred to the concept of a house, resulting in only two classes: NDI-body, and NDI-house.

Verbs, on the other hand, required more postprocessing. VIIs showed the best clustering results without postprocessing. For example, VII_6 was entirely made up of taste/smell lexemes, VII_7 verbs were almost entirely weather-related, VII_8 contained verbs that only take plural subjects (the semantic nature of which is discussed below), VII_9 had only lexemes referring to sound and sight, and VII_{10} had only nominal-like verbs (e.g. *mîsiyâpiskâw* '(it is) rust(y)').[9] Despite these well-formed clusters, VII_1 through VII_5 were less cohesive and required manual clustering. In the end, 6 distinct classes were identified: II-natural-land, II-weather[10], II-sensory, II-collective[11], II-move, II-named[12]. Although postprocessing was required, this was not too substantial in scope or time. The VAIs required significantly more work. Some classes were well defined, such as VAI_23 whose members all described some sort of flight, but VAI_{12} contains verbs of expectoration, singing, dancing, and even painting. Rather than being able to consolidate some classes, most HAC-produced classes needed to be manually split further. Although here one could have cut the HAC tree at a lower level to create more classes. This did not produce better or cohesive classes. The resulting VAI classes were as follows: AI-state, AI-action, AI-reflexive, AI-cooking, AI-speech, AI-collective, AI-care, AI-heat/fire, AI-money/count, AI-pray, AI-childcare, AI-canine[13], and AI-cover. The VTIs similarly required manual postprocessing after HAC clustering. Although some classes

[9]Although this form may be thought of as attributive, an identical form is used as an NI. Whether this is a separate lexeme, or a nominal use of a verb is debatable.

[10]This class includes terms of weather as well as terms of seasons or times of day such as sîkwan, 'it is spring'.

[11]The semantic status of this class is discussed below.

[12]This class contains terms of being named, such as *isiyihkâcikâtêw*, 'It is named thus'

[13]This class refers to verbs that specifically describes behaviors specific to canines, e.g. nêmow, 's/he growls as a dog'.

such as VTI_{11} (entirely to do with cutting or breaking) or VTI_{14} (entirely to do with pulling) were very well formed, the majority of the classes needed further subdivision (though significantly less so than with the VAIs), resulting in the following 6 classes: TI-action, TI-nonaction, TI-speech, TI-money/count, TI-fit, and TI-food. Finally, the VTAs required a similar amount of postprocessing as the VAIs. Although a few classes were well formed (such as VTA_4 which was entirely made up of verbs for 'causing' something), the vast majority of HAC classes contained two or more clear semantic groupings. Through manual postprocessing, the following set of classes were defined: TA-action, TA-nonaction, TA-speech, TA-food, TA-money/count, TA-religion, and TA-allow.

3.4.1 Evaluation

In addition to the qualitative evaluation presented above, I present a preliminary quantitative evaluation of this technique. This evaluation allows us to judge how useful these classes are in practical terms, providing an indirect measure of the informational value of the clusters. I make use of the mixed effects modelling that initially motivated automatic semantic clustering, focusing on a morphological alternation called Nêhiyawêwin Order, wherein a verb may take the form *ninipân* (the *Independent*) or *ê-nipâyân* (the *ê-Conjunct*), both of which may be translated as 'I sleep.' The exact details of this alternation remain unclear, though there appears to be some syntactic and pragmatic motivation (Cook, 2014). Using R (R Core Team, 2017) and the lme4 package (Bates et al., 2015), I ran a logistic regression to predict alternation using verbal semantic classes as categorical variables. In order to isolate the effect of semantic class, no other effects were used. The semantic classes were included as random effects. To assess the effectiveness of semantic class in this context, I assessed the pseudo-R^2 value, a measure of Goodness-of-Fit. Unlike a regular R^2 measure, the pseudo-R^2 can not be interpreted as a direct measure of how much a model explains variance, and generally "good" pseudo-R^2

Table 3.2: pseudo-R^2 Values for Modelling Independent vs. ê-Conjunct Order Choice Based on Manual and Automatic Clustering Evaluation. Larger values represent better model fits.

	Manual	HAC-Only
VII	0.18	0.19
VAI	0.13	0.09
VTI	0.04	0.01
VTA	0.06	0.06

value are comparatively smaller (McFadden, 1973), though a higher value still represents a better fit. As a general rule, a pseudo-R^2 of 0.20 to 0.40 represents a well fit model (McFadden, 1977).[14] Models were fit for each of the four conjugation classes for both classes produced directly from the Hierarchical Agglomerative Clustering as well those manually adjusted. I used a subset of the Ahenakew-Wolfart Corpus (Arppe et al., 2020), containing 10,764 verb tokens observed in either the Independent or ê-Conjunct forms. The resulting pseudo-R^2 scores represent the way in which automatic and semi-manual clusters can explain the Nêhiyawêwin Order alternation.

Table 3.2 presents the result of these analyses. the *Manual* column represents clusters that were manually adjusted, while the *HAC-Only* column represents the result of the logistic model that used only the fully automatic HAC-produced clusters. A larger value in the table represent a model that is better able to predict the Order a verb takes. While this prediction is not in and of itself of primary importance to the focus of this section, a better fitting model in this context suggests the efficacy of fully-manual vs. semi-automatic verb classification. If, for example, the semi-automatic classification scheme produced a less explanatory model than the fully-automatic scheme, there would be no reason to spend the time and effort for the semi-manual classification task. Further, if either model neglected to show any significant explanatory power, one would have no reason to include semantic classes in future predictive models at all.

The manually adjusted and HAC-only classes performed similarly, especially for

[14]One can also compare the results in this paper with results from a similar alternation study in Arppe (2008).

VTAs, though manual adjustment had a slightly worse fit for the VIIs, and conversely the VAI and VTI have somewhat significantly better fits using the manually adjusted classes. Although it appears that manual adjustment produced classes that were somewhat better able to explain this alternation, both manually adjusted and HAC-only clusters appear to explain a non-negligible degree of this alternation phenomenon in the above models. This is significant, because it shows that the result of the clustering techniques presented in this paper produce a tangible and useful product for linguistic analysis. Further, it suggests that, although manual classification was sometimes more useful, automatic classes more or less performed as well, allowing for researchers to determine if the added effort is worth the small increase in informational value. Nevertheless, alternative methods of evaluation, such as evaluating clusters based on speaker input, particularly through visual means as described in Majewska et al. (2020) should be considered.[15]

3.5 Discussion

In general, the best clustering was seen in classes with fewer items. The VAI and NI lexemes required the most postprocessing, with each having roughly double the number of items as the next most numerous verb/noun class. Verb classes in general seemed to produce less cohesive classes through HAC. Although the exact cause of this discrepancy in unknown, it could perhaps be due to the way words are defined in Wolvengrey (2001). In this dictionary, verb definitions almost always contain more words than noun definitions. Almost every single verb definition will have at least two words, owing to the fact that Nêhiyawêwin verbs are defined by an inflected lexeme. This means that if one looks up a word like *walk*, it would appear as: pimohtêw: s/he walks, s/he walks along; s/he goes along. Meanwhile, nouns tend to have shorter definitions. The definition for the act of walking, a nominalized form of the verb for walk, is written as: pimohtêwin:

[15]It is worth noting that previous attempts at such experimentation via Nêhiyawêwin communities with which I have good relationships have been poorly received by speakers.

walk, stroll; sidewalk. This difference is exacerbated by the fact that definitions are often translated fairly literally. Something like *pêyakwêyimisow* might be translated simply as 's/he is selfish,' but contains morphemes meaning *one*, *think*, *reflexive*, and *s/he*. A gloss of this word is seen in (11). Rather than simply defining the word as 's/he is selfish,' Wolvengrey (2001) has opted to provide a more nuanced definition: pêyakwêyimisow: s/he thinks only of him/herself, s/he is selfish, s/he is self-centered.

(11) *pêyakwêyimisow*
pêyakw-êyi-m-iso-w
one-think-VTA-RFLX-3SG

's/he thinks only of him/herself'

The result of this complex form of defining is that words are defined more in line with how they are understood within the Nêhiyawêwin culture, which is indeed often manifested in the derivational morphological composition of these words. This is central to the motivation for this method of semi-automatic clustering, but produces verbs with relatively long definitions. An alternative explanation for why Nêhiyawêwin lexemes with English definitions consisting of more numerous parts of speech were more difficult to classify is that these divisions simply have significantly more variation in meaning for whatever reason. Further investigation into this is needed.

Also worth noting is the relative distributions of each of the postprocessed classes mentioned above. Table 3.3 details each of the postprocessed noun classes sorted by their size.

Perhaps unsurprisingly, the distribution of lexemes into different classes followed a sort of Zipfian distribution. The NA-person and NA-beast-of-burden accounted for the vast majority of noun lexemes for animate nouns. Just under half of all NI lexemes were nominalized verbs, and roughly a quarter were smaller object-like items (e.g. tools, dishes, etc.). The NDAs were almost entirely dominated by words for family, while all but three NDIs were body part lexemes. Some categories such as NI-scent, NI-days, and

Table 3.3: Manually Adjusted Noun Classes

NI (N)	NDI (N)	NA (N)	NDA (N)
NI-nominal (1783)	NDI-body (243)	NA-persons (720)	NDA-relations (143)
NI-object (902)	NDI-house (2)	NA-beast-of-burden (512)	NDA-body (45)
NI-natural-Force (283)		NA-food (325)	NDA-clothing (4)
NI-place (228)		NA-celestial (45)	
NI-nature-plants (198)		NA-body-part (37)	
NI-body-part (78)		NA-religion (23)	
NI-hunt-trap (60)		NA-money/count (12)	
NI-animal-product (48)		NA-shield (2)	
NI-religion (36)			
NI-alteration (23)			
NI-scent (4)			
NI-days (4)			
NI-persons (3)			

NA-shield have extremely low membership counts, but were substantially different from other categories that they were not grouped into another class. Most interestingly, there appeared to be three NI lexemes that referred to persons, something usually reserved for NAs only. These lexemes were *okitahamâkêw* 'one who forbids,' *owiyasiwêwikimâw* 'magistrate,' and *mihkokwayawêw* 'red neck.' In all three cases, the lexemes seem to be deverbal nouns (from *kitahamâkêw* 's/he forbids,' *wiyasiwêw* 's/he makes laws,' and *mihkokwayawêw* 's/he has a red neck.'

Verbs showed a similar distribution. Table 3.4 details the distribution of words within each of the semantic classes for verbs. With the exception of VII and VAIs, verbs were dominated by classes for action (as these are transiutive classes, this is unsurprising), which subsumes most volitional actions (e.g. *kîskihkwêpisiwêw* 's/he rips the face off of people,' *kâsîpayiw* 's/he deletes'), and nonaction which includes most verbs of thought, emotion, judgment, or sensory action (e.g *koskowihêw*, 's/he startles someone,' *nôcîhkawêw* 's/he seduces someone'). Other classes may include action verbs, such as AI-cooking and TI-speech. Although these verbs could be classified in one of the two previously mentioned systems, their automatic classification and semantics unify them in a way that is unique to other items in these larger classes.

Table 3.4: Manually Adjusted Verb Classes

VII (N)	VAI (N)	VTI (N)	VTA (N)
II-natural-land (275)	AI-state (2083)	TI-action (1409)	TA-action (1013)
II-weather (96)	AI-action (1982)	TI-nonaction (293)	TA-nonaction (574)
II-sensory (90)	AI-reflexive (542)	TI-speech (80)	TA-speech (103)
II-collective (79)	AI-cooking (172)	TI-money/count	TA-food (54)
II-move (38)	AI-speech (131)	TI-fit (10)	TA-money/count (23)
II-named (3)	AI-collective (97)	TI-food (8)	TA-religion (9)
	AI-care (81)		TA-allow (5)
	AI-heat/fire (55)		
	AI-money/count (34)		
	AI-pray (29)		
	AI-childcare (17)		
	AI-canine (16)		
	AI-cover (15)		

Verbs in AI-action have little in common with each other except that they are a form of volitional action, while AI-care verbs (which may include actions related to giving care such as *kanawastimwêw* 's/he looks after/guards horses') have a distinct and unifying characteristic relating to giving care. Similarly, although AI-childcare could be subsumed under AI-care, the former includes items like *kimotôsêw* 's/he bears an illegitimate child'. This is even more obvious in categories such as AI-collective or AI-reflexive, which refer to lexemes that are plural only or reflexive in nature/morphology, respectively. These may not seem as semantically defined as other classes for VAIs, though one could argue that verbs that occur only in the plural are inherently collective in action, and thus semantically defined; similarly, reflexive forms are necessarily actions that are done to one's self. Although there may be action or nonaction verbs in this category, the automatic classification divided and grouped most reflexive and plural only lexemes into their own respective classes. Resultantly, these clusters were kept as separate classes. For this classification scheme, reflexives were deemed to be more reflexive than they were to be action or nonaction.

Additionally, VAIs contained a sort of stative class, AI-state. This classification, being inherently non-transitive, is not present in the VTI or VTA classes. Stative verbs

are present in the VII class, but given how many VII lexemes are essentially stative, I opted not to have a single stative class, but instead defined classes describing natural-land (including general landscape features such as *kinohtakâw* 'it has a long floor'), sensory information (e.g. *kihcinâkwan* 'it looks impressive'), or weather terms (*mispon* 'it is snowing').

Overall, verb forms, especially the most numerous classes of VAI and VTA, required a large degree of manual postprocessing. Because this approach assumes no underlying ontology, but rather attempts to work bottom-up (cf. Hanks 1996), the time taken to postprocess VAI and VTA classes is likely not too far from what it would take to manually classify these words based off a prebuilt ontology; however, the appeal of a bottom-up classification should not be overlooked.

3.6 Conclusion

This paper describes an attempt at semi-automatically classifying Nêhiyawêwin verbs and nouns. Resulting clusters of Nêhiyawêwin words are freely available online. Although the technique worked quite well with nouns, which required very little manual adjustment, verbs required more directed attention. Despite this, the technique presented in this paper offers a bottom-up, data-driven approach that takes the language on its own terms, without resorting to ontologies created primarily for other languages. If, however, one wishes to use a pre-defined ontology, the basis for this work (representing word definitions using pre-trained English word vectors) could be used in conjunction with existing ontologies to expedite the classification process. For example, Dacanay et al. (2021) compare the naive definition vectors for Wolvengrey (2001) with the same for the English WordNet word senses; word senses whose vectors bear a strong correlation with the Nêhiyawêwin definitions can then be assumed to be synonymous with a Nêhiyawêwin word, and the latter can take the WordNet classification of the former. Because this technique leverages

resources from a majority language, it is not sensitive to the issue of paucity of data for minority languages. It should be applicable to any context where a minority language has a majority-language-bilingual-dictionary and where the majority language is well resourced. Applications for this research extend not only to the creation of semantic classes, but also to the association of words based on semantic similarity. The results of the quantitative evaluation suggest that, at least in the Independent vs. ê-Conjunct alternation, semantic class plays some role in predicting the alternation, though its use varied by conjugation class. In addition to the use of these results to bolster modelling of Nêhiyawêwin Order, the word similarity scores on which clustering was based can be used to identify words that are similar to one another, a task that is ideal for word discovery, for example in the presentation of synonymous (or at least semantically related) terms when searching through an online dictionary.

Future research should investigate how these classes compare to raw HAC clusters and manual classification of various sorts (should these become available in Nêhiyawêwin). Different methods of calculating item distance in clustering techniques (e.g. through cosine distance (Dacanay et al., 2021)) should be considered. More sophisticated sentence/definition embeddings, such as those returned by BERT (Devlin et al., 2018) or other state of the art models would also likely increase the efficacy of this technique. Further, as one reviewer suggested, one could use a weighted average for words in the dictionary definitions along with word relevance measures (such as tf-idf scores) to more accurately represent the semantics of an English sentence. Although fully Nêhiyawêwin-trained vectors are ideal, as with most Indigenous languages of North America, there is simply nowhere close to enough data to build robust word embeddings as seen in the Google News Corpus. The technique described in this paper presents a compromise of taking the language on its own terms, while leveraging the massive data sets that exist for majority languages.

Chapter 4

Methodology

This chapter details the methods used in the analysis of Nêhiyawêwin Order. The primary research question investigated in this dissertation is: how, and in what way, can Nêhiyawêwin order be understood as an alternation that can be predicted through morphosyntactic, surface-syntactic, and lexical-semantic features. This chapter describes the corpus used, the univariate analysis, and the multivariate analysis. The methodologies used in this analysis are based off those univariate, bivariate, and multivariate statistics described in Bresnan et al. (2007); Divjak (2010); Gries (2003)[1] and Arppe (2008), in particular the combination of univariate and multivariate techniques. This chapter only generally covers the methods used for creating the underlying corpus, which has been described at length in Arppe et al. (2020). The process by which verbs and nouns were semantically clustered for inclusion as predictors is also not detailed in this section (as it is described in Chapter 3).

[1]Though Gries (2003) uses discriminant analysis instead of regression, his general multi-level analysis framework is followed here, as in Arppe (2008).

4.1 The Corpus

The underlying corpus from which the data set used in this dissertation is the Ahenakew-Wolfart corpus (Arppe et al., 2020). The Ahenakew-Wolfart corpus is likely the largest morphosyntactically tagged corpus of all Canadian Indigenous languages, let alone Nêhiyawêwin. Although there have been attempts in the last few decades to increase the amount of texts in Nêhiyawêwin through the publishing of stories as lectures originally composed in the language (as in Ahenakew (2000), Masuskapoe (2010), etc.), there is still a paucal amount of texts available, and those that are, are written in a nonstandard Roman orthography. The Ahenakew-Wolfart corpus is unique in that it is meticulously standardized. The texts that make up the corpus were collected, transcribed, and translated by Freda Ahenakew and H. C. Wolfart between the 1970s and 1990s. These texts have previously been published in Ahenakew (2000); Bear et al. (1992); Kâ-Nîpitêhtêw (1998); Masuskapoe (2010); Minde (1997); Vandall and Douquette (1987); Whitecalf (1993). These texts are mainly dialectic or narrative discussions between two or more native Nêhiyawêwin speakers. Together, these texts contain 142,192 tokens (20,503 types), though some of these tokens are English, French, or Michif words; fragments; or other items. Focusing only on Nêhiyawêwin items, there are 80,221 tokens (16,532 types). Each of these tokens has been morphosyntactically tagged by automatic and hand-parsed means (Arppe et al., 2020), based on an initial pass by a finite-state morphological analyzer (Harrigan et al., 2017; Snoek et al., 2014). Tokens were lemmatized as well as tagged for morphosyntactic features. For verbs, these concern preverbs, tense, word class, Order, comitative morphemes, and conjugation class; for nouns, these concern possession/number marking, possession, declension, and diminutive morphemes. Both nouns and verbs were marked for the feature of semantic class. An example token with its relevant tags is found in (12) and (13).

(12) ê-ohci-pimâtisit
pimâtisiw PV/e PV/ohci V AI Cnj 3Sg @PRED-AI

'S/he lived thus / made a living thus'

(13) kikâwînaw
nikâwiy N A D Px21Pl Sg @ACTOR>

'Our Mother/the Earth'

Beyond this, the corpus has been further disambiguated and syntactically tagged by a constraint grammar parser (Schmirler, 2022; Schmirler et al., 2018). Among other features, this parser marks tokens for their status as predicate, actor, or goal.

To create the data set used in this dissertation, I extracted only verbs from the above corpus and further restricted the data set by selecting only verbs that contained a semantic classification as described in Chapter 3. This resulted in a data set of 13,628 tokens (2,032 types). In addition to the morphosyntactic tags seen above, verbs were marked for arguments (and those arguments' morphosyntactic features) when arguments were syntactically present (as opposed to represented only by person marking on the verb). This resulted in an entries such as (14).

(14) ê-ohci-pimâtisit pimâtisiw PV/e PV/ohci V AI Cnj 3Sg @PRED-AI AI-state

kikâwînaw nikâwîy N A D Px1Sg Sg @ACTOR> NDA-Relations

From here, each token and its accompanying analyses were transformed into a data frame of logical variables: every verb lemma token makes up a row, while every morphosyntactic or semantic tag constitutes a logical column. For every lemma token, if a feature is observed, a value of `TRUE` is set for the corresponding column, otherwise a value of `FALSE` is set. Logical variables allow for easily interpreted results, especially when dealing with covariance (Arppe, 2008; Baayen, 2012). Given the example of (14), the data frame extract in Table 4.1 is produced.

Table 4.1: Extract from Data Frame

Lemma	PRED-AI	PV/ahci	...	PV/e	PV/ohci	PV/pe	V	AI	Cnj	3.actor	3.goal	AI-state	Sg.actor
pimâtisiw	TRUE	FALSE	...	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE

For the sake of fitting Table 4.1 to the page, the majority of the columns are not shown, but every feature present in (14) would have a column value of `TRUE` for the token `pimâtisiw`, and all features not present are given a value of `FALSE`. The exception to this is the actor and goal marking morphemes. Although the corpus marks person and number of morpheme as one unit (e.g. `1Sg`), the data frame used for analysis in this dissertation splits the features up (i.e. there were separate columns for `3` and `Sg` for both actors and goals). Finally, a number of tokens were removed because their verb conjugation class was not reliably identified in the corpus. There were 310 of these tokens, the majority of which (301 tokens) were the verb *ayâw*. In addition to basic locative use, *ayâw* may also be used to describe the state of 'having' something. In the corpus, *ayâw* was marked as both `VAI` and `VTI`. Because the VTI form of the verb inflects the same as the VAI form, and because syntactic arguments are usually not present in a sentence beyond verbal agreement (and even then, only in the VTA), determining which of these two classes the lemma was acting in was difficult for the non-native speakers annotating the corpus. Three further lemmas, *manitowi-kîsikâw* (4 tokens), *misi-paskwâw* (3 tokens), and *nanamipayiw* (1 token), were removed as there was disagreement between the corpus and dictionary sources. In the first two cases, these forms were given in the corpus as VIIs, while dictionary sources cited them also as NIs. This disagreement is understandable, as VIIs that deal with time or space often describe substantives. The final case, *nanamapayiw* is given as an VII in the corpus, while Wolvengrey (2001) analyses it as an VAI and LeClaire et al. (1998) offers an analysis of both VAI and VII. Although context (through either native speaker annotation or translations by native speakers) would quickly resolve these ambiguities, the corpus being used had not yet been disambiguated in this sense, and given the small number of

Table 4.2: Preverb Class Tokens and Types

	Types	Tokens
Discourse	4	277
Position	15	285
Qual	30	316
Quant	7	10
Time	18	4720
Move	4	731
Start/Finish	5	229
Want/Can	4	195

tokens, I opted to remove these 309 tokens from the data set.

Because Nêhiyawêwin contains a large number of possible preverbs (the model underlying the corpus could identify 267 unique preverbs), I undertook a manual classification of these morphemes.[2] I identified 8 unique classes: Discourse ,[3] Position , Qual , Quant , Time , Move , Start/Finish and Want/Can . Of the 267 identified preverbs, only 86 preverb types were observed in the corpus. Table 4.2 lists the number of tokens and types in each of the preverb classes.

In all, the resulting data frame of non-imperative forms contains 13,292 lemma rows by 4777 columns. Due to errors in coding (e.g. a number of items were misidentified as nouns when they were verbs), 100 items were excluded from this, creating a data frame of 13,192 items. The use of such a logical data frame for predicting an alternation is presented in Arppe (2008) and allows for the assessment of individual values of categorical variables through straightforward application of chi-squared analyses and logistic regression to predict a multinomial[4] alternation, in this case Order.

[2]This process entailed inspecting the definition of all preverbs and grouping together those morphs with high level, abstract similarites. For example, while *miyo-* ('good') and *mâyi-* ('bad') had opposite meanings, they each conceptually represented qualitative valuation of some thing. As a result, they were grouped together under the Qual category.

[3]Preverbs of Discourse include {isi-} ('so, thus, such'), {aya-} ('um') and similar preverbs that do not provide direct content in an utterance, do not indicate information such as tense or aspect, but which nonetheless function to build the utterance/discourse.

[4]The term *multinomial* is used here instead of the term *polytomous* as in Arppe (2008). I use this term (except when referring to the specific package by Arppe (2013)) as the term makes specific reference to

Table 4.3: Preverb Class Tokens and Types

		Types	Tokens
Independent		876	4390
ê-Conjunct		1480	6378
kâ-Conjunct		600	1696
Other-Conjunct		393	828
	Subjunctive	75	100
	Initial Change	18	21
	ka-Conjunct	344	707
Total		3349	13,294

4.2 Modelling the Alternations

In this dissertation, I will evaluate a univariate analysis given the morphosyntactic and semantic features mentioned above to model a verb lemma's likelihood of occurring in various Order types. Although Chapter 2 identified five unique Conjunct Orders (along with the Independent), the majority of these classes have few tokens. Small counts can be problematic for statistical analyses, particularly for regression. To address this, the ka-/ta-Initial, Initial Change, and Subjunctive Conjuncts were conglomerated into a single 'other' class. This results in the Order alternations as seen in Table 4.3.

In order to gain a wholistic understanding of Nêhiyawêwin Order, this dissertation will investigate 3 main alternations of these Orders:

- Independent vs. Conjunct
- Independent vs. ê-Conjunct
- Conjunct Type: ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct

The first of these alternations, Independent vs. Conjunct, will inform about the difference between the two Orders broadly. The second, Independent vs. ê-Conjunct,

nominal (not ordered) data, which is the sort of data of primary interest to this dissertation.

Table 4.4: `AWnImp` *statistics*

	Types	Tokens
Independent	876	4390
Conjunct	1722	8802
Total	2598	13,192

Table 4.5: `AWIvE` *statistics*

	Types	Tokens
Independent	876	4390
ê-Conjunct	1480	6378
Total	2356	10,768

will investigate the difference between the two most similar Order forms which are often conceived as synonymous and used roughly interchangeably. The third alternation will be used to model the extent to which we can predict the modes through morphosemantic features from a corpus. Three main data frames were used:

- `AWnImp` : used in analyzing the Independent vs. Conjunct alternation, representing all non-imperative forms minus the 100 errors previously mentioned
- `AWIvE` : used in analyzing the Independent vs. ê-Conjunct alternation, representing only forms with ê-Conjunct and Independent forms
- `AWCnj` : used in analyzing the Conjunct Type alternation, representing only forms with Conjunct forms (of any kind).

In Table (4.4) through (4.6) are relevant counts for each of the three dataframes.

Table 4.6: AWCnj *statistics*

	Types	Tokens
ê-Conjunct	1480	6378
kâ-Conjunct	600	1696
Other-Conjunct	393	828
Total	2473	8902

4.3 Univariate Analyses

The term *univariate analysis* refers to an analysis that takes into account only one explanatory variable at a time while explaining the occurrence of some other dependent outcome variable. The most common form of univariate analysis for discrete variables is the chi-square test, originally introduced in Pearson (1900) and refined over the last century to produce the modern day chi-squared test (Agresti, 2013). The chi-square test makes use of contingency tables to measure the association/correlation of a (set of) values of one (explanatory) categorical value against the values of an outcome variable. This is calculated by comparing the expected frequency of an outcome/variable pair with with the observed frequencies of the same pairings. Chi-square tests provide a simple statistic, the eponymous χ^2 statistic, whose value reflects an estimated association. This statistic is given for the whole *set* of values of the explanatory and outcome variables tested. If one were to run a chi-square test to determine if the set of variables { 1Sg.actor , 2sg.actor , 3sg.actor , past tense , future tense , present tense } was associated with an increased likelihood of a lemma being in the Independent or Conjunct Order, the resulting χ^2 statistic would indicate the level of association for that set as a whole. To investigate the effect an individual variable has, one must make use of the Standardized Pearson Residuals (SPR), calculated through the formula in (4.1), where *P* is the Standardized Pearson Residual, *O* is the observed frequency of a variable/outcome pair, *E* is the Expected frequency of a variable/outcome pair, t_i is the sum of a variable across all

outcomes, and t_j is the sum of all variables for a given outcome (adapted from (Agresti, 2013, 81)). Note that in (4.1) the denominator represents its standard error.

$$P = \frac{O - E}{\sqrt{E(1 - t_i)(1 - t_j)}} \tag{4.1}$$

This produces Standardized Pearson Residuals which can be interpreted based on its magnitude and direction. A positive SPR of at least 2.00 represents a significant positive association (i.e. one observes more instances of a variable/outcome pairing than would be expected) while a negative value of -2.00 or lower represents a negative association. Values greater than -2.00 but less than 2.00 represent an association not deemed to be significant (Agresti 2013, 81; exemplified in Arppe 2008, 79).

The chi-square test is best used with higher frequency data sets. According to Cochran (1954), the results of a chi-square test are not reliable when the contingency tables for a given variable has more than 20% of its expected values <5. In these cases, it is suggested that researchers make use of an alternative test, such as the Fisher's Exact Test that forms the basis of Gries' Collostructional Analysis (2004). Some authors, however, believe that Fisher's Exact Test is too conservative (D'Agostino et al., 1988), increasing the risk for Type II errors in hypothesis testing. For this dissertation, I will simply consider phenomena with sufficient frequencies for a chi-square statistic.

Univariate Models

In building models for univariate analysis, all variables with a minimum occurrence of 10 were selected for a given conjugation class for each alternation. This restriction was chosen to exclude incredibly infrequent items which make statistical modelling difficult or unreliable, while including as many variables as possible. Because univariate analysis considers variables on their own basis, manual scrutiny of variable selection was not performed at this point.

4.4 Bivariate Analyses

Following Arppe (2008), after univariate analyses were conducted and a set of variables were selected, I conducted bivariate analyses. Bivariate analysis is simply measuring the association between two variables. In effect, bivariate analysis is a special case of univariate analysis, where one contrasts two explanatory variables against each other, rather than one explanatory variable against the outcome variable. Bivariate analysis as done by Arppe (2008) can be a useful tool for creating models for mixed effects modelling. Bivariate analysis for this dissertation makes use of the `associations` function from the `polytmous` package (Arppe, 2013). This function calculates Theil's Uncertainty Coefficient (Henri, 1970) for every combination of variables passed to it. This coefficient is a mutual information measure and describes the extent to which knowing about one variable can inform our understanding of another variable via a reduction of uncertainty (Arppe, 2008, 90).

4.4.1 Bivariate models

Bivariance was tested for each of the four alternations mentioned above. Variables for each alternation were chosen only from those items with a significant χ^2 statistic ($p < 0.05$). Automatic and manual classes were tested separately, as there was a great deal bivariance between automatic and manual semantic class variables.

4.5 Multivariate Analysis

Using the methodology of Arppe (2008), following bivariate analysis, the resulting variable sets were used to form a set of variables to perform multivariate analysis. The fundamental technique used in this analysis was logistic mixed effects regression. Logistic regression is a generalized form of linear regression as applied to categorical

outcomes. Logistic regression models a binary outcome based on a (set of) predictor variables(s) (Agresti, 2013, 163).

Logistic regression (with a single independent variable, for example) can be modelled with the equation in (4.2) (Agresti, 2013, 163), where x represents the independent variable, β is the slope of x (the extent to which x effects an outcome), and α represents a model intercept (the chance of an outcome occurring when there is no predictor).

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \tag{4.2}$$

The value resulting from (4.2) represent the odds ratio for the effect of an independent variable on a particular outcome (e.g. the effect of register on the use of one of two synonyms). Here, the odds ratio represents the probability of an outcome ocurring divided by the probability of it not occurring. These ratios are bounded between 0 and ∞. More commonly, however, logistic regression models are fit with the logit function, derived from (4.2) and seen in (4.3) (Agresti, 2013, 163).

$$\mathrm{logit}[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x \tag{4.3}$$

The resulting estimates given by the logit are given in log-odds, rather than odds. These values are *not* bounded between 0 and ∞, but instead $-\infty$ and $+\infty$. Positive values represent an increase in likelihood of an outcome for a particular variable; negative values represent a decrease in likelihood; a value of zero represents no effect on the outcome.

Like all generalized linear models then, logistic regression attempts to predict outcomes by representing the distribution of the data. Specifically, the technique allows researchers to specify a set of predictors and models the data so they can determine the extent to which an individual predictor influences a particular outcome given a set of parameters (variables/effects).

This dissertation makes use of mixed effects models in its logistic regression. In terms

of regression for language data, mixed effects models have now become the norm (Barth and Kapatsinski, 2018, 100). In comparison to models that make use of only fixed effects, those variables for which all possible values are represented in the data, mixed effects models allow for the researcher to control for variables in which random variation can be expected (Baayen, 2012). For the data used in this dissertation, morphosyntactic features like `Actor.1`, which are dummy variables that represent the presence or absence of a feature (in this case, whether or not a verb is marked for first person), are **fixed effects** because all possible values (`TRUE, FALSE`) are represented in the data. Conversely, the `Lemma` variable (a multi-level variable containing all lemmas of the corpus) are *samples* of the total lemma set in Nêhiyawêwein and thus can be expected to contain some amount of random variability/outcomes not present in the corpus; thus, `Lemma` is best modelled as a **random effect**. In a mixed effects model, the random variability of a random effect is 'controlled' for, allowing for estimations of fixed effects without the confounds of the random effects.

Fixed effects are analyzed relatively straightforwardly: for each of the logical variables, one of two possible classes are chosen to act as a baseline reference (Baayen, 2012). This baseline acts as a reference point for analyses: if the *presence* of a variable is chosen as the baseline, then what is measured by observation is the opposite; that is, whether or not the variable did **not** occur. The programming language used in this dissertation to statistically analyze results, R (R Core Team, 2017), by default uses the `0/FALSE` level as a base line, though one could use the alternate level as a reference if needed. For the logical variables in this dissertation, the reference level represents the *absence* of a particular variable. In modelling an outcome, the logistic regression analyzes each observation in its training data and, if an outcome is *not* observed, assigns the variable a value of 0 for the outcome; otherwise, if the variable *is* observed, a value of 1 is given to the variable for the outcome (Baayen, 2012). Importantly, a model's intercept represents all of the variables' reference levels (Baayen, 2012). Random effects

are not given a reference level; instead, each level can be thought of as adjustments to each fixed effect (Baayen, 2012). As an example, given the fixed effect `actor.1`, the logistic model would make adjustments to `actor.1`'s slope based on observations of each level of `Lemma`. In this sense, there is no reference level the others are compared to.

This mixed-effects analysis makes use of the `lme4` package in R (Bates et al., 2015).

4.5.1 Binarization of the Alternation

Making use of logistic regression, this dissertation will investigate the behaviour of three alternations. The comparisons made allow for investigation of a wide range of Order behaviour.

Logistic regression assumes a dichotomous decision by default. This is the case, for example, when comparing the Independent and the Conjunct. For the final two alternations above, however, there are more than two outcomes being compared. In multinomial cases, there are multiple methods by which the data can be binarized. One such technique is the *one-vs-rest* (OVR) heuristic. In one-vs-rest comparisons, a model compares one class against all other possible classes (Frank and Kramer, 2004). In this way it can be thought of as comparing x against ¬x. One can also make use of the *pairwise comparison*, which assesses all possible pairings of some set of variables to determine the likeliest outcome. In the present study, each Conjunct class would be paired with one other class at a time, eventually being paired with each other Conjunct type. For each pairing, which-ever type is most probable given each set of predictors, would be picked as the 'winner' for that pairing, and would be proposed to the model. For each morphological feature, that class which is most often proposed would be selected and given as the likeliest Order type. In addition to one-vs-rest and pairwise comparisons, we can make use of *nested dichotomies*. The concept of a nested dichotomy is straightforwardly described: a group of outcomes being compared is split

into mutually exclusive dichotomies repeatedly until unary classifications are created (Frank and Kramer, 2004). Each dichotomy is given a statistical probability which can be multiplied together to determine the overall probability of the dichotomous tree. Similarly, other binarization techniques such as Baseline-Category classification (Fox, 1997, 468), wherein a single outcome of a multinomial decision is arbitrarily selected as a baseline/default/prototypical case against which other outcomes are compared, offer alternative techniques. Arppe (2008) presents a detailed overview of all the above techniques as they relate to linguistic analysis; while Frank and Kramer (2004) presents a detailed overview of the general case of such methods. For this dissertation, I will use the OVR heuristic, as Arppe (2008) found useful, in the multinomial Conjunct Type alternation due to its conceptual and computational ease and simplicity (that is, the direct calculation of log-odds that affect each outcome), particularly in modelling maximally three outcomes in a single alternation.

The resulting logistic models provide estimated effect magnitudes for every variable, even if these effects are not considered statistically significant. The results provided by the `lme4` package calculates the p-value for each effect using the asymptotic Wald tests for generalized linear models (Bates et al., 2015). Recognizing that the use of p-values are not without controversy (Gelman, 2016), this dissertation will still use p-values to determine which effects are most pertinent in modelling the Order alternation.

4.6 Model Assessment

In addition to the results described above, one can assess the overall performance of a logistic model. This assessment gives us invaluable information and allows us to see how well, and in what ways, a model represents Order type selection in terms of morphosyntactic and semantic features. In particular, one can see how well a model is able to evaluate a given form as the correct Order type without raising false

positives (precision), as well as how many instances of a given Order type it classifies correctly, regardless of false positives (recall). Recall, precision, and overall accuracy are measured per Conjunct type for each conjugation class (not for the model in its entirety). As an additional way of assessing model fit, one may also use the τ value to determine how much better the model performs than selecting based solely on overall proportions/through random assignment (Goodman and Kruskal, 1959, 745-747) (ranging from 0.00 to 1.00, with 1.00 being a perfect model). According to Arppe (2008, 140), τ values of roughly 0.5 and above suggest a 'good' model fit. Because the models used in this dissertation are logistic, a true R^2 score assessing estimated variance is inapplicable. Instead, a so-called Pseudo-R^2 value must be used. As Hosmer and Lemeshow (2000, 167) point out, Pseudo-R^2 Likelihood scores for logistic regression are generally much smaller than in other statistics, such as R^2 values given in standard linear models. Another important difference between the R^2 measure and Pseudo-R^2 Likelihood is that the former can be used as a measure of how much variance is explained by the model under consideration; Pseudo-R^2 Likelihood does not report explained variance (Hosmer and Lemeshow, 2000, 164). Instead, Pseudo-R^2 Likelihood can be seen as a measure of reduction in the *badness* of fit. The specific form of Pseudo-R^2 that I will use is McFadden's Pseudo-R^2 (ρ^2) (Domencich and McFadden, 1975) as reported by the `ModelStatistics` function (Arppe, 2013). Macfadden's Pseudo-R^2 appears to have a stable, but non-linear, relationship with a general R^2, wherein a ρ^2 value of 0.2, 0.3, and 0.4 are roughly equivalent to an R^2 of 0.3, 0.5, and 0.73 respectively (Domencich and McFadden, 1975, 124). As with other Pseudo-R^2 measures, a ρ^2 of over 0.25 is indicative of a fairly well fit model. Further, Han et al. (2013) suggest that, in their experience, Pseudo-R^2 Likelihood scores of nearly 0.30 are indicative of very good models without risk of over-fitting.

Although the results of the logistic regression are given in log-odds, this is straightforwardly transformed to probability by reversing the process of (4.2). We can

do this by exponentiation of the log-odds we derived, as in (4.4):

$$p = \frac{e^x}{1 + e^x} \tag{4.4}$$

That is, we derive the estimated probability of an outcome by raising e to the power of our log-odds (x) and dividing this by $1 + e$ to the power of our log-odds. The result is a simple estimated probability of an outcome resulting, given a (set of) predictor(s).

Because models are fit separately for each Conjunct type in each conjugation class in the Conjunct Type alternation, estimated probabilities for a set of variables add up to something close to, but not exactly, 1.00. In order to achieve this range of 0.00-1.00, the `ModelStatistics` function (Arppe, 2013) aggregates all models and performs a normalization of estimated probabilities, such that they add up to exactly one.

In addition to the models described above, one can use logistic regression with only random effects. The models constructed here present only those lemma-specific effects, and can be used as a base-line against which one can assess how much of an effect morphological features have on the ability to predict Conjunct type (cf. the discussion of Harrigan and Arppe (2015) regarding lemma-specific preferences on occurrence in the Independent or Conjunct orders). For each of the mixed-effect models, Pseudo-R^2 Likelihood, Accuracy, and τ measures for models with only random effects will be also be given. By comparing fixed effects models against the mixed effects models, we can determine the extent to which random effects affect the fit of our modelling of Order.

4.6.1 Working Predictions

Based on the above methodology, the following predictions are proposed:

1. Overall, modelling will be successful though constrained (likely due to the small size of the corpus).

2. Due to a lack of syntactic (at least in terms of detailed information about phrase

structure, c-command, adjunct/complement and other syntactic criteria, mixed effects modelling based on the Nêhiyawêwin corpus will be able to provide some insights, but model fits will rarely achieve a level beyond well fitting.

3. Semantic variables will do more to explain variance than morphosyntactic variables (as in (Arppe, 2008)).

4. It will be difficult to humanly aggregate significant results of the Conjunct Type Alternation into a useful interpretation) in its results (due to the straightforward syntactic/semantic choices driving it, which are not reflected in the variables of the data set).

5. The alternation between the Independent and the ê-Conjunct will be the most well-fit model (because the two forms are nearly synonymous in many cases).

Chapter 5

Results

Results presented in this chapter represent only the statistically significant results of modelling. Full results can be found at https://github.com/atticusha/DissertationCode. Where *p* values are given, the values are given to two decimal places unless doing so would result in a rounded value of 0.00. In this case, the value is reported to three decimal places. If the *p* value is less than 0.001, the result is reported as <0.001. Some results *appear* non-significant (i.e. having a reported *p* value of 0.05 or higher) due to rounding; however, these values are below 0.05 before rounding for presentation.

5.1 Univariate results

Three separate univariate analyses were undertaken, one for each of the alternations being studied. Generally speaking, these analayses identify those variables having a statistically significant association with a particular alternation outcome.

The following three subsections detail the models used for univariate analysis as well as the resulting (significant) variables. Variables fed into the univariate models followed the selection criteria described in Chapter 4, but certain adjustments were made. Given their low counts, variables indicating specific argument lemmas (e.g. aya.goal) were

removed. Instead, I used the more abstract semantic classes from Chapter 3 in their place. Additionally, corpus-internal tags (like `Lemma` and `Morph`, which indicate a lemma or morpheme not present in the semi-automated gold-standard corpus) were not included in analyses. Similarly, tags indicating the direction of the argument in relation to the verb (e.g. `@ACTOR>.actor` for arguments occurring to the left of the verb) were not used. Although it is possible some syntactic information *could* be helpful, Nêhiyawêwin word order is very flexible, and only slightly less than half of all verbs even contained overt arguments. Further, previous syntactic accounts have not suggested linear order of arguments to be a significant influence on Order. Finally, tags like `N.actor` (which represents that an actor was a syntactically instantiated noun) and `I.actor` (which represents a syntactically instantiated inanimate noun) were removed because they were implicitly reflected in the verb class (e.g. all actors are nouns, all VAIs will have animate actors).[1]

5.1.1 Independent vs. Conjunct

The selected variables were fed into the `nominal` R function, producing a set of χ^2 test results. Statistically significant results are presented in Tables 5.1 through 5.4. These tables depict the predictor names, the number of tokens for each predictor, their χ^2 statistic, and direction of association between predictor and outcome (i.e. a `+` indicates a significant positive association, which in turn implies that a predictor occurs signficantly more often with a particular alternation construction than would be expected by chance, and a `-` represents the opposite; a `0` represents no signficang association).

Intransitive Inanimate Verbs

For the Independent vs. Conjunct alternation VIIs, only preverbs of time and sensory verbs were positively associated with the Independent Order. All other effects such

[1] Inanimate actor forms are not included in this corpus

Table 5.1: Significant univariate results for the Independent Vs. Conjunct Alternation: VIIs

	N	χ^2	Cnj	IND
PV.Time	213	<0.001	-	+
II.sense	274	<0.001	-	+
NI.object.actor	144	0.026	+	-
Pron.actor	58	0.02	+	-
Dem.actor	57	0.01	+	-
Med.actor	24	0.03	+	-

as *Inanimate Object Actor*, as well as pronominal, demonstrative, and medial actors, were positively associated with the Conjunct. Note that both *Demonstrative Actors* and *Medial Actors* are features associated with certain pronouns, and so are essentially subtypes of *Pronominal Actors*. As well, because *Medial Actors* represents medial demonstrative pronouns, all *Medial Actors* forms are implicitly also *Demonstrative Actors*.

Intransitive Animate Verb

Table 5.2: Significant univariate results for the Independent Vs. Conjunct Alternation: VAIs

	N	χ^2	CNJ	IND
AI.speech	1344	<0.001	-	+
actor.3	3646	<0.001	-	+
PV.Time	2008	<0.001	+	-
PV.Move	442	<0.001	+	-
PV.Qual	173	0.003	+	-
PV.StartFin	143	0.003	+	-
PV.Discourse	137	<0.001	+	-
PV.Position	121	0.004	+	-
AI.do	2125	<0.001	+	-
AI.state	1924	<0.001	+	-
AI.cooking	281	<0.001	+	-
AI.reflexive	276	<0.001	+	-
AI.health	122	0.02	+	-
AI.pray	61	0.005	+	-
RdplW	142	<0.001	+	-
NA.persons.actor	737	<0.001	+	-
Sg.actor	540	0.001	+	-
Pl.actor	295	<0.001	+	-
Pron.actor	403	0.01	+	-
Dem.actor	200	0.02	+	-
NA.beast.of.burden.actor	59	0.02	+	-
NA.food.actor	37	0.03	+	-
actor.1	1825	0.04	+	-
actor.2	250	<0.001	+	-
Obv.actor	45	0.01	+	-

The VAIs showed a number of significant effects, nearly all positively associated with the Conjunct as detailed in Table 5.2. In fact, only verbs relating to speech or those having a third person actor seemed to positively associate with the Independent. Notably, a number of preverbs were positively associated with the Conjunct: those of time, movement, quality, starting/finishing, discourse, and position.[2] In general, it appears that semantic effects were associated with the Conjunct, with the semantic classes of action, state, cooking, reflexive, heath, and praying all having significant positive Conjunct effects. Beyond these and the effect of weak/light reduplication, all other remaining

[2]In fact, the category of 'positional' preverbs, are far less likely to do with literal, spatial position, and are overwhelmingly the preverb {ohci-} used metaphorically as a negative past marker. This is discussed in more detail in Chapter 6. When terms like 'preverbs of position' or 'position preverbs,' are used, this fact is understood. The term 'position' in this dissertation is used as a convention, rather than in a literal sense.

effects were related to explicitly realized actors (as separate words). Singular and plural actors, pronominal and demonstrative, actors semantically relating to food and beasts of burden, as well as first, second, and obviative actors were all positively associated with the Conjunct.

Transitive Inanimate Verbs

Table 5.3: Significant univariate results for the Independent Vs. Conjunct Alternation: VTIs

	N	χ^2	CNJ	IND
TI.cognitive	1160	<0.001	-	+
NA.persons.actor	261	0.02	-	+
Pron.actor	158	<0.001	-	+
Pers.actor[3]	107	<0.001	-	+
actor.1	1202	<0.001	-	+
actor.2	250	<0.001	-	+
Px1Sg.goal	18	0.05	-	+
TI.do	1632	<0.001	+	-
TI.money.count	23	0.03	+	-
PV.Discourse	64	<0.001	+	-
NI.nominal.goal	114	0.002	+	-
NI.natural.force.goal	73	0.03	+	-
NI.place.goal	42	0.001	+	-
Sg.goal	789	<0.001	+	-
Pl.goal	244	<0.001	+	-
D.goal	64	0.01	+	-
NDI.Body.goal	55	0.01	+	-
Px3Sg.goal	43	0.02	+	-
Der.Dim.goal	30	0.01	+	-
actor.3	1514	<0.001	+	-

Like the VAIs, there were a number of VTI effects that showed a significant association, mostly with the Conjunct, as seen in Table 5.3. The only Independent-associated effects (though the accounted of roughly one third of tokens) were verbs having to do with cognition, actors that had referents that were types of people, pronominal actors (especially personal pronouns), first or second person actors, and goals possessed by singular first persons. The majority of Conjunct associated variables concerned arguments, specifically goals. The only verbal associations were the semantic classes of

[3]This variable represents the presence of personal pronouns as actors.

action and money/counting as well as preverbs of discourse. Goals that were nominalized verbs, natural forces or place names; singular or plural goals, dependent goals, dependent goals specifically relating to body parts, those possessed by a singular third persons, and diminutive goals all associated with the Conjunct. The only actor based effect for the Conjunct was that of third persons actors, the opposite of the VTAs.

Transitive Animate Verbs

Table 5.4: Significant univariate results for the Independent Vs. Conjunct Alternation: VTAs

	N	χ^2	CNJ	IND
TA.speech	1114	<0.001	-	+
goal.3	1498	<0.001	-	+
actor.1	1067		-	+
PV.Time	1189	<0.001	+	-
PV.Move	172	<0.001	+	-
PV.Discourse	68	<0.001	+	-
PV.Qual	62	0.05	+	-
PV.Position	47	0.01	+	-
TA.cognitive	843	0.02	+	-
TA.do	837	<0.001	+	-
TA.food	96	<0.001	+	-
TA.money.count	66	0.04	+	-
goal.obv	702	<0.001	+	-
goal.2	177	0.002	+	-
NA.persons.goal	394	0.001	+	-
Px3Pl.goal	20	0.03	+	-
NA.persons.actor	199	0.01	+	-
NDA.Relations.actor	82	0.02	+	-
Sg.actor	178	<0.001	+	-
D.actor	82	0.02	+	-
actor.3	1266	<0.001	+	-
actor.obv	152	<0.001	+	-

The VTAs followed a similar pattern as seen previously. Nearly all significant effects in Table 5.4 were positively associated with the Conjunct, though verbs having to do with speech, having a first person actor, and having a third person goal all positively associated with the Independent. As in the VAIs, all preverb effects (those of time, movement, discourse, quality, and position) were associated with the Conjunct. Verbs of cognition, action, food, and money/counting were similarly aligned. Both actor and goal effects were

present in the VTAs. Second person and obviative goals, those possessed by plural third persons, and goals representing people were all positively associated with the Conjunct. Actor effects such as the semantic classes of person actors and those representing a dependent relationship, singular and dependent actors, as well as third and obviative persons were also positively associated with the Conjunct.

5.1.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Table 5.5: Significant univariate results for the Independent Vs. ê-Conjunct Alternation: VIIs

	N	χ^2	ê-Cnj	Ind
II.sense	255	0.001	-	+
PV.Time	186	<0.001	-	+
NI.object.actor	126	0.001	+	-
Sg.actor	158	0.03	+	-
Pron.actor	48	0.007	+	-
Dem.actor	47	0.003	+	-
Med.actor	21	0.01	+	-

In the Independent vs. ê-Conjunct, VIIs demonstrated a positive association between sensory verbs and preverbs of time with the Independent outcome. Inanimate object, singular, pronoun, demonstrative, and medial actors were all positively associated with the ê-Conjunct.

Intransitive Animate Verbs

Table 5.6: Significant univariate results for the Independent Vs. ê-Conjunct Alternation: VAIs

	N	χ^2	ê-Cnj	Ind
AI.speech	1240	<0.001	-	+
actor.3	3109	<0.001	-	+
PV.Time	1654	<0.001	+	-
PV.Move	346	<0.001	+	-
PV.Qual	147	0.001	+	-
PV.StartFin	123	<0.001	+	-
PV.Discourse	116	<0.001	+	-
PV.Position	109	<0.001	+	-
AI.do	1671	<0.001	+	-
AI.state	1578	<0.001	+	-
AI.cooking	231	<0.001	+	-
AI.reflexive	222	<0.001	+	-
AI.pray	46	0.012	+	-
RdplW	124	<0.001	+	-
NA.persons.actor	554	0.004	+	-
NA.beast.of.burden.actor	46	0.03	+	-
Pl.actor	228	0.01	+	-
actor.1	1551	0.002	+	-
actor.obv	138	<0.001	+	-

The VAIs continued previous trends: verbs of speech with third person actors associated with the Independent, but all other significant effects as described in Table 5.6 were positively associated with the ê-Conjunct. This includes the verbal effects: preverbs of time, movement, quality, starting/finishing, discourse, and position; semantic classes, of verbs of action, state, cooking, praying, and reflexive verbs; and weak/light reduplication all positively associated with the ê-Conjunct. Actor effects included person actors, beasts of burden, plural actors, first person actors, and obviative actors.

Transitive Inanimate Verbs

Table 5.7: Significant univariate results for the Independent Vs. ê-Conjunct Alternation: VTIs

	N	χ^2	ê-Cnj	Ind
TI.cognitive	1008	<0.001	-	+
NA.persons.actor	203	0.003	-	+
Pers.actor	95	<0.001	-	+
Pron.actor	129	<0.001	-	+
Px1Sg.goal	15	0.05	-	+
actor.2	181	<0.001	-	+
actor.1	1043	<0.001	-	+
TI.do	1281	<0.001	+	-
TI.money.count	17	0.04	+	-
PV.Discourse	55	<0.001	+	-
NI.natural.force.goal	64	0.008	+	-
NI.place.goal	31	0.002	+	-
Der.Dim.goal	20	0.02	+	-
Pl.goal	205	<0.001	+	-
actor.3	1184	<0.001	+	-

The VTIs showed a more equal distribution for Independent and Conjunct effects. Verbs of cognition, actors representing people, pronominal actors (especially personal pronouns), actors which are possessed by first persons, and verbs with first and second person actors positively associated with the Independent. Conversely, verbs of action, verbs of money/counting, preverbs of discourse, goals representing natural forces, goals representing places, and plural goals all positively associated with the ê-Conjunct. The only actor-based positive ê-Conjunct association was third person actors.

Transitive Animate Verbs

Table 5.8: Significant univariate results for the Independent Vs. ê-Conjunct Alternation: VTAs

	N	χ^2	ê-Cnj	Ind
TA.speech	905	<0.001	-	+
actor.1	892	<0.001	-	+
actor.2	84	<0.001	-	+
goal.3	1185	<0.001	-	+
PV.Time	946	<0.001	+	-
PV.Move	123	<0.001	+	-
PV.Discourse	58	<0.001	+	-
PV.Qual	49	0.04	+	-
PV.Position	44	0.001	+	-
PV.StartFin	30	0.03	+	-
TA.cognitive	692	0.003	+	-
TA.do	650	<0.001	+	-
TA.food	80	<0.001	+	-
Sg.actor	138	0.001	+	-
D.actor	68	0.009	+	-
NDA.Relations.actor	68	0.009	+	-
Obv.actor	31	0.02	+	-
actor.3	1060	<0.001	+	-
actor.obv	109	<0.001	+	-
goal.obv	579	<0.001	+	-
NA.persons.goal	290	0.04	+	-
Px3Sg.goal	36	0.01	+	-
Px3Pl.goal	15	0.03	+	-

The VTAs in the Independent vs. ê-Conjunct alternation mostly exhibit significant associations with the ê-Conjunct: only verbs of speech, local actors, and third person goals showed an association with the Independent Order. The usual significant preverb classes, those of time, movement, discourse, quality, position and starting/finishing were associated with the ê-Conjunct, as were the major semantic classes of cognition, action, and food. A number of actor effects positively associated with the ê-Conjunct, including singular and dependent actors, actors representing dependent relations, and non-local actors. Goals which were obviative, representative of persons, and those that were possessed by third persons were also associated with the outcome.

5.1.3 Conjunct Types

Unlike the previous two sections, the alternation described in this section is multinomial. As a result, the positive/negative association for one outcome does not imply the opposite association in another outcome. Some items may show a 0 mark in the tables representing the lack of a significant effect in any particular outcome.

Intransitive Inanimate Verbs

Table 5.9: Significant univariate results for the Conjunct Type Alternation: VIIs

	N	χ^2	ê-Cnj	kâ-Cnj	Other-Cnj
II.sense	171	<0.001	+	-	0
PV.Time	135	0.001	+	-	0
Sg.actor	134	<0.001	+	-	0
NI.object.actor	119	<0.001	+	-	0
II.natural.land	145	<0.001	-	+	0

The VIIs showed significant associations only for the ê- and kâ-Conjuncts, where the two always showed association in the opposite direction (a pattern which can be partially seen throughout this alternation). Sensory verbs, preverbs of time, singular arctors, and object actors were positively associated with the ê-Conjunct and negatively associated with the kâ-Conjunct. Verbs representing nature/land were the odd ones out, positively associating with kâ-Conjunct and negatively associating with the ê-Conjunct.

Intransitive Animate Verbs

Table 5.10: Significant univariate results for the Conjunct Type Alternation: VAIs

	N	χ^2	ê-Cnj	kâ-Cnj	Other-Cnj
PV.Discourse	120	0.05	+	-	0.00
PV.Position	95	0.01	+	-	0.00
actor.1	1251	0.001	+	-	0.00
actor.3	2222	<0.001	+	0.00	-
RdplW	114	0.03	+	0.00	0.00
actor.2	207	<0.001	-	+	+
AI.do	1649	0.005	-	+	0.00
NA.persons.actor	545	<0.001	-	+	0.00
Sg.actor	392	<0.001	-	+	0.00
Pron.actor	292	<0.001	-	+	0.00
Prox.actor	91	<0.001	-	+	0.00
Dem.actor	149	<0.001	-	+	-
AI.health	99	0.001	-	0.00	+
Pl.actor	224	<0.001	0.00	+	-
Med.actor	56	0.05	0.00	+	0.00
PV.Time	1442	<0.001	0.00	-	+
PV.WantCan	45	<0.001	0.00	-	+
AI.cooking	222	0.03	0.00	-	0.00
PV.Qual	133	0.02	0.00	0.00	-
D.actor	164	0.003	0.00	0.00	-
NDA.Relations.actor	164	0.003	0.00	0.00	-
Px1Sg.actor	123	0.02	0.00	0.00	-

The VAIs showed a more varied result. Preverbs of discourse and position, as well as first person actors were positively associated with the ê-Conjunct and negatively associated with the kâ-Conjunct. Third person actors were positively associated with the ê-Conjunct and negatively associated with the Other-Conjunct. The final positive association for the ê-Conjunct was weak/light duplication, which was only significant for the ê-Conjunct. Verbs of action, as well as actors that were people, signular, pronouns, proximate, or demonstrative all positively associated with the kâ-Conjunct while negatively associating with the ê-Conjunct. Plural actors and medial actors were positively associated with the kâ-Conjunct, while preverbs of time, desire/ability and verbs of cooking were negatively associated with the outcome. Interestingly, the Other-Conjunct outcome regularly disagreed in association with the ê-Conjunct: third person actors had a positive association with the ê-Conjunct, but they had a negative association

with the Other-Conjunct; while the opposite pattern is seen with second person actors and verbs of health. Other positive associations with the Other-Conjunct outcome were preverbs of time (likely a result of the ka-Conjunct necessarily having PV.ka present, a preverb of time) and desire/ability. The class had negative effects in the form of demonstrative, plural, and dependent actors, preverbs of quality, actors representing dependent relations, and actors possessed by a singular first person.

Transitive Inanimate Verbs

Table 5.11: Significant univariate results for the Conjunct Type Alternation: VTIs

	N	χ^2	ê-Cnj	kâ-Cnj	Other-Cnj
PV.Position	45	0.002	+	-	-
actor.1	693	<0.001	+	0.00	-
PV.WantCan	51	<0.001	-	+	0.00
TI.speech	171	<0.001	-	+	0.00
NA.persons.actor	162	<0.001	-	+	0.00
Sg.goal	586	<0.001	-	+	+
NI.object.goal	512	<0.001	-	+	0.00
Dem.goal	216	<0.001	-	+	0.00
Pron.goal	216	<0.001	-	+	0.00
Prox.goal	122	<0.001	-	+	0.00
NI.nominal.goal	95	<0.001	-	+	+
actor.2	137	<0.001	-	0.00	+
Med.goal	94	0.05	-	0.00	0.00
D.goal	54	0.05	-	0.00	+
PV.Time	860	<0.001	0.00	-	+
Sg.actor	93	0.01	0.00	+	0.00
Pron.actor	88	0.05	0.00	+	0.00
actor.3	1195	0.05	0.00	0.00	-

The VTIs had only two significant positive associations for the ê-Conjunct: position preverbs and first person actors. Second person actors and those which were people, along with singular, object, nominal, demonstrative, pronominal, proximal, medial and dependent goals all negatively associated with this outcome, as did verbs of speech and preverbs of desire/ability. Conversely, only two forms showed a negative association with the kâ-Conjunct: position preverbs and time. All other significant effects, including preverbs of desire/ability; singular, pronominal, and person actors; verbs of speech;

and goals representing objects, nominalized verbs, singular entities, demonstratives, pronouns, and proximals all occurred more often with the kâ-Conjunct than otherwise would be expected based on chance. position preverbs and first or third person actors were negatively associated with the Other-Conjunct, while singular, nominal, and dependent goals were positively associated with the outcome. Beyond these, second person actors and preverbs of time were also positively associated with the Other-Conjunct.

Transitive Animate Verbs

Table 5.12: Significant univariate results for the Conjunct Type Alternation: VTAs

	N	χ^2	ê-Cnj	kâ-Cnj	Other-Cnj
PV.Time	863	<0.001	+	-	+
TA.cognitive	580	<0.001	+	-	+
actor.3	893	<0.001	+	-	-
goal.obv	529	<0.001	+	-	-
PV.Position	39	0.01	+	-	0.00
goal.1	392	<0.001	+	0.00	-
PV.Discourse	58	0.03	+	0.00	-
TA.speech	576	<0.001	-	+	0.00
Sg.goal	195	<0.001	-	+	0.00
Prox.goal	66	0.02	-	+	0.00
actor.2	84	<0.001	-	0.00	+
goal.3	897	<0.001	-	0.00	+
goal.2	134	0.01	-	0.00	+
Prox.actor	31	0.01	0.00	+	0.00
Px1Sg.goal	67	0.04	0.00	0.00	-
actor.1	567	0.03	0.00	0.00	-

Results for the VTAs present a number of different significant effects for all outcomes. Preverbs of time, position, and discourse; third person actors, first person and obviative goals, and verbs of cognition all positively associated with the ê-Conjunct. Second and third person goals, proximate goals, singular goals, second person actors, and verbs of speech were all negatively associated with the outcome. The kâ-Conjunct was positively associated with proximate actors, proximate and singular goals, and verbs of speech. Preverbs of time and position, third person actors, obviative goals, and verbs of

cognition were negatively associated with the outcome. Finally, preverbs of time, verbs of cognition, second person actor and goals, and third person goals all positively associated with the Other-Conjunct outcome; first person actors, third person actors, first person and obviative goals, and prevebs of discourse all negatively associated with the Other-Conjunct.

5.2 Bivariate Results

Before moving on to the multivariate analysis, the significant effects from the univariate analysis for each verb class in each alternation were tested for pairwise association. Pairs which were found to be bivariate (those with Theil's uncertainty coefficients (Theil, 1970) of > 0.50 (as in Arppe (2008), indicating that one the presence of absence of one variable provides information about the presence of absence of another or reduces uncertainty by at least 50%) were dealt with by removing one of the items in the pair. This is done as, according to Harrell (2017, 64), variables that can predict another variable, result in large standard errors, and thus the statistical power of the model is reduced. The item for removal was chosen based on its relevance and explanatory value, e.g. if *Dependent Goals* and *Inanimate, Dependent, Body Part Goals* are bivariate, the latter provides more semantic information than the former, and so the former would be removed.

The following section presents the effects that formed a bivariate pair along with the eventual list of effects to address this bivariance. The resulting list of effects will be used for multivariate analysis in the next subsection.

The bivariance results present the effects that form each bivariate pair (referred to as *category1* and *category2*), as well as the number of tokens for each category (e.g. *N1* is the number of tokens for *category1*, *N2* is the number of tokens for *category2*). The resulting tables also include the number of tokens where each effect co-occur, represented in column *N12*, and the uncertainty coeffecients (where *uc.12* gives the percent by which

Category 1 predicts the presence or absence of *Category 2* and *uc.21* gives the inverse.).

5.2.1 Independent vs. Conjunct

Intransitive Inanimate Verbs

Table 5.13: Bivariate results for the Independent vs. Conjunct Alternation: VIIs

category1	category2	N1	N2	N12	uc.12	uc.21
NI.object.actor	Pron.actor	144	58	58	0.29	0.53
NI.object.actor	Dem.actor	144	57	57	0.29	0.53
Pron.actor	Dem.actor	58	57	57	0.96	0.98
Pron.actor	Med.actor	58	24	24	0.33	0.63
Dem.actor	Med.actor	57	24	24	0.34	0.64

Bivariance in the VIIs concerned actor variables. Every instance of both *Pronominal Actors* and *Demonstrative Goals* were used along with the *Inanimate Object Actors* tag. This confirms that when demonstrative pronouns are used with VIIs as actors, they represent inanimate objects. Similarly, the bivariate results in Table 5.13 also show that nearly all pronominal actors were demonstrative and that all demonstrative pronouns were medial; less interestingly, all medial pronouns coocurred with pronominal tags. To alleviate bivariance, `Pron.actor` and `Dem.actor` were removed, resulting in the following variables to be kept for multivariate analysis: `PV.Time` , `II.sense` , `NI.object.actor` , `Med.actor` .

Intransitive Animate Verbs

Table 5.14: Bivariate results for the Independent vs. Conjunct Alternation: VAIs

category1	category2	N1	N2	N12	uc.12	uc.21
NA.persons.actor	Pron.actor	737	405	405	0.44	0.66
NA.persons.actor	Dem.actor	737	201	201	0.40	0.52
Pron.actor	Dem.actor	403	201	201	0.41	0.69
actor.3	actor.1	3646	1836	0.00	0.49	0.56

Again, the VAI's bivariance for the Independent vs. Conjunct focused on actor variables. Whenever a demonstrative or more general pronoun was observed so too was an actor from the NA.persons.actor class. As a consequence of demonstrative pronouns being pronouns, the two were similarly bivariate. Finally, third and first person actors were bivariate, as the latter never cooccured with the former. Pron.actor , Dem.actor , and actor.1 were removed to address bivariance, resulting in the following variables for multivariate analysis: PV.Time , PV.Move , PV.Qual , PV.StartFin , PV.Discourse , PV.Position , AI.do , AI.state , AI.speech , AI.cooking , AI.reflexive , AI.health , AI.pray , RdplW , NA.persons.actor , Sg.actor , Pl.actor , NA.beast.of.burden.actor , NA.food.actor , actor.3 , actor.2 , actor.4 .

Transitive Inanimate Verbs

Table 5.15: Bivariate results for the Independent vs. Conjunct Alternation: VTIs

category1	category2	N1	N2	N12	uc.12	uc.21
TI.do	TI.cognitive	1632	1163	0.00	0.64	0.66
NA.persons.actor	Pron.actor	261	158	158	0.50	0.72
NA.persons.actor	Pers.actor	261	107	107	0.32	0.62
Pron.actor	Pers.actor	158	107	107	0.59	0.79
actor.3	actor.1	1508	1202	0.00	0.58	0.60
D.goal	NDI.Body.goal	64	55	55	0.80	0.91

The VTIs show a similar pattern as the above classes: Pronouns, and specifically personal pronouns, always occured with actors representing people, and personal pronouns were necessarily also pronouns. Similarly, all dependent goals having to do with body parts were also classified as dependent nouns. As in the VAI class, first and third person actors never cooccured. Finally, the two main verb classes, TI.do and TI.cognitive were also bivariate, never occurring together. To address this bivariance, TI.cognitive , Pron.actor , Pers.actor , actor.1 , and D.goal were removed leaving the following variables: PV.Discourse , TI.do , TI.money.count , NA.persons.actor , actor.3 ,

actor.2 , Sg.goal , Pl.goal , NI.nominal.goal , NI.natural.force.goal , NDI.Body.goal , Px3Sg.goal , NI.place.goal , Der.Dim.goal , Px1Sg.goal .

Transitive Animate Verbs

Table 5.16: Bivariate results for the Independent vs. Conjunct Alternation: VTAs

category1	category2	N1	N2	N12	uc.12	uc.21
D.actor	NDA.Relations.actor	82	82	82	1.00	1.00
actor.3	goal.3	1266	1498	0.00	0.68	0.66

The VTAs had a much smaller set of covariates than previous classes. All dependent actors representing people of close relation were also marked as dependent goals (for obvious reasons). Third person actors and goals were also bivariate, never occurring together (as one argument would need to be obviative in terms of Nêhiyawêwin grammar). Here, D.actor and actor.3 , were removed, leaving the following variables for multivariate analysis: PV.Time , PV.Move , PV.Discourse , PV.Qual , PV.Position , TA.speech , TA.cognitive , TA.do , TA.food , TA.money.count , NA.persons.actor , Sg.actor , NDA.Relations.actor , actor.1 , actor.4 , goal.3 , goal.obv , goal.2 , NA.persons.goal , Px3Pl.goal .

5.2.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Table 5.17: Bivariate results for the Independent vs. Conjunct Alternation: VIIs

category1	category2	N1	N2	N12	uc.12	uc.21
Sg.actor	NI.object.actor	158	126	119	0.54	0.60
NI.object.actor	Pron.actor	126	48	48	0.27	0.50
Pron.actor	Dem.actor	48	47	47	0.96	0.97
Pron.actor	Med.actor	48	21	21	0.35	0.64
Dem.actor	Med.actor	47	21	21	0.36	0.65

Once again, the VII effects concerned only actors. Nearly every instance of NI.object.actor also occurred with Sg.actor. In turn, the Pron.actor tag always occurred with NI.object.actor while Dem.actor always occured with Pron.actor. Similarly, Med.actor always occurred with both Pron.actor and Dem.actor. Removing Sg.actor, Pron.actor, and Med.actor leaves PV.Time, II.sense, NI.object.actor, and Dem.actor as variables.

Intransitive Animate Verbs

There were no bivariate pairs with substantial uncertainty scores for VAIs in the Independent vs. ê-Conjunct alternation.

Transitive Inanimate Verbs

Table 5.18: Bivariate results for the Independent vs. Conjunct Alternation: VTIs

category1	category2	N1	N2	N12	uc.12	uc.21
TI.do	TI.cognitive	1281	1008	0.00	0.68	0.69
NA.persons.actor	Pron.actor	203	129	129	0.53	0.74
NA.persons.actor	Pers.actor	203	95	95	0.37	0.65
Pron.actor	Pers.actor	129	95	95	0.65	0.82
actor.3	actor.1	1184	1043	0.00	0.62	0.63

As in the previous alternation, verbs of cognition and verbs of action were bivariate in that they never occur together. Similar to other classes, Pron.actor and Pers.actor always occured with NA.persons.actor, and Pers.actor did the same with Pron.actor. Finally, third and first person actors never cooccured. To alleviate this covariance, effects of TI.Cognitive, Pron.actor, Pers.actor, and actor.1 were removed, leaving PV.Discourse, TI.do, TI.money.count, NA.persons.actor, actor.3, actor.2, Pl.goal, NI.natural.force.goal, NI.place.goal, Der.Dim.goal, Px1Sg.goal as the final set of variables for multivariate analysis.

Transitive Animate Verbs

Table 5.19: Bivariate results for the Independent vs. Conjunct Alternation: VTAs

category1	category2	N1	N2	N12	uc.12	uc.21
D.actor	NDA.Relations.actor	68	68	68	1.00	1.00
actor.3	goal.3	1060	1185	0.00	0.69	0.69

For VTAs. there were only two instances of bivariance: NDA.Relations.actor always occurred with D.actor, and third person actors and goals never occurred together. D.actor and actor.3 were removed to produce the following set of variables: PV.Time, PV.Move, PV.Discourse, PV.Qual, PV.Position, PV.StartFin, TA.speech, TA.cognitive, TA.do, TA.food, Sg.actor, D.actor, NDA.Relations.actor, actor.1, actor.4, actor.2, goal.3, goal.obv, NA.persons.goal, Px3Sg.goal, and Px3Pl.goal

5.2.3 Conjunct Type

Intransitive Inanimate Verbs

Table 5.20: Bivariate results for the ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct Alternation: VIIs

category1	category2	N1	N2	N12	uc.12	uc.21
Sg.actor	NI.object.actor	134	119	110	0.60	0.64

The single instance of bivariance for the VIIs in the Conjunct Type alternation was the relationship between Sg.actor and NI.object.actor, where the latter nearly always occured alongside the former. Removing the Sg.actor produce the variable set: PV.Time, II.sense, II.natural.land, II.weather, NI.object.actor.

Intransitive Animate Verbs

Table 5.21: Bivariate results for the ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct Alternation: VAIs

category1	category2	N1	N2	N12	uc.12	uc.21
NA.persons.actor	Pron.actor	545	292	292	0.42	0.65
NA.persons.actor	Dem.actor	545	149	149	0.20	0.50
Pron.actor	Dem.actor	292	149	149	0.42	0.69
Pron.actor	Prox.actor	292	91	91	0.24	0.59
Pron.actor	Med.actor	292	56	56	0.15	0.52
D.actor	NDA.Relations.actor	164	164	164	1.00	1.00
D.actor	Px1Sg.actor	164	123	123	0.67	0.83
NDA.Relations.actor	Px1Sg.actor	164	123	123	0.67	0.83
Dem.actor	Prox.actor	149	91	91	0.55	0.77
Dem.actor	Med.actor	149	56	56	0.31	0.67

In the VAIs, Pron.actor and Dem.actor always cooccured with NA.persons.actor ; both Dem.actor , Prox.actor , and Med.actor were always accompanied by Pron.actor ; NDA.Relations.actor and Px1Sg.actor always coccured with D.actor ; Px1Sg.actor was always accompanied by NDA.Relations.actor . Finally, both Prox.actor and Med.actor always cooccured with Dem.actor . Removing Prox.actor , Pron.actor , Dem.actor , Med.actor , D.actor , Px1Sg.actor results in the following variables: PV.Time , PV.Qual , PV.Discourse , PV.Position , PV.WantCan , AI.do , AI.cooking , AI.health , RdplW , NA.persons.actor , Sg.actor , Pl.actor , NDA.Relations.actor , actor.3 , actor.1 , actor.2 .

Transitive Inanimate Verbs

Table 5.22: Bivariate results for the ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct Alternation: VTIs

category1	category2	N1	N2	N12	uc.12	uc.21
NA.persons.actor	Pron.actor	162	88	88 & 0.44	0.70	
actor.3	actor.1	1195	693	0.00	0.58	0.62
Dem.goal	Pron.goal	216	216	216	1.00	1.00
Dem.goal	Prox.goal	216	122	122	0.46	0.68
Dem.goal	Med.goal	216	94	94	0.34	0.62
Pron.goal	Prox.goal	216	122	122	0.46	0.68
Pron.goal	Med.goal	216	94	94	0.34	0.62

The bivariance for VTIs in the Conjunct Type Alternation concerned mostly goal related variables. The variables Pron.goal , Prox.goal , and Med.goal always cooccured with Dem.goal . Both Prox.goal and Med.goal cooccured with Pron.goal . Removing Pron.actor , actor.1 , Dem.goal , Pron.goal results in: PV.Time , PV.WantCan , PV.Position , TI.speech , NA.persons.actor , Sg.actor , actor.3 , actor.2 , Sg.goal , NI.object.goal , Prox.goal , NI.nominal.goal , Med.goal , D.goal .

Transitive Animate Verbs

Table 5.23: Bivariate results for the ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct Alternation: VTAs

category1	category2	N1	N2	N12	uc.12	uc.21
actor.3	goal.3	893	897	0.00	0.65	0.65

The final class to discuss is the VTAs, which had only a single bivariate pair. In this pair, actor.3 and goal.3 never occurred together. Thus, these two features perfectly predict one another and one must be removed to deal with exact co-linearity. Removing Actor.3 results in the final set of variables: PV.Time , PV.Discourse , PV.Position ,

TA.cognitive , TA.speech , Prox.actor , actor.1 , actor.2 , goal.3 , goal.obv , goal.1 , goal.2 , Sg.goal , Px1Sg.goal , Prox.goal .

5.3 Multivariate results

The following section details the results of the multivariate logistic regressions described in Chapter 4. These results are presented as a set of tables where each row represents a fixed effect (i.e. those effects identified in the previous section). In all cases, results include the inclusion of lemma type as a random effect. In addition, each table contains a row labelled *Intercept*. Like the effects, the intercept is not reported if non-significant (though it is available in the supplementary repository[4] for this dissertation).The intercept represents the effect for the aggregate of all the implicit values that are excluded from the set of variables used in modeling. As Agresti points out, the intercept is not usually of much explanatory value in and of itself (Agresti, 2013, 165), though to calculate probability esitmates, it is necessary. Each effect is reported with an estimate of impact (in log-odds) of the associated effect, as well as an *p*-value. A summary table is given for each of the four verb classes in each of the three alternations being studied.

5.3.1 Independent vs. Conjunct

In this alternation, all effects are reported for their influence on the occurrence of an Independent form of a verb (as contrasted with a Conjunct form). If an effect is positive, a verb is more likely to occur in the Independent with it and less likely to do so for the Conjunct. On the other hand, if an effect is negative, a verb is less likely to occur in the Independent, and more likely to do so in the Conjunct.

[4]https://github.com/atticusha/DissertationCode/

Intransitive Inanimate Verbs

Table 5.24: Multivariate results for the Independent vs. Conjunct Alternation: VIIs

	Independent	
	Estimate	*p*-value
(Intercept)	-1.370	< 0.001
PV.Time	0.673	0.001

In the alternation between the Independent and the Conjunct generally, the VIIs had only a single significant effect: preverbs of time, which increase the likelihood of an Independent form.[5] In fact, of the 204 Independent forms, 81 contained a preverb of time, the vast majority of which (57) were the past tense PV.ki, as in (15) and (16).

(15) “êy, **kî-miywâsin**,” itwê-w, nôcikwêsiw ana ...
Hey, PST-be.good.3.SG say.3.SG, old.woman that ...

‘ “Hey, life used to be good” she said, that old woman ...’ (Bear et al., 1992, 74)

(16) **kî-âyiman** ôtê ka-pê-wîcihiwê-yân maskwacîsihk ...
PST-be.difficult.3.SG here PST-come-live-1.SG maskwacîsihk ...

‘it was hard to come live here at maskwacîsihk ...’ (Minde, 1997, 2)

In each of the above examples, the verbs represent matrix clause verbs, particularly in (16) where an embedded verb, *ka-pê-wîcihiwêyân*, appears in a Conjunct form. This characterization of the Independent as a matrix form and the Conjunct as an embedded form fits with the description of the Order in Cook (2014).

[5]This may be related to the fact that that tense in Independent clauses is absolute and thus operates without reference to an antecedent, while tense in Conjunct clauses is relative (Wolvengrey, 2012). Despite this, a similar effect was not seen throughout the alternation.

Intransitive Animate Verbs

Table 5.25: Multivariate results for the Independent vs. Conjunct Alternation: VAIs

	Independent	
	Estimate	*p*-value
(Intercept)	-1.379	< 0.001
PV.Discourse	-0.944	0.001
actor.obv	-0.812	0.001
Sg.actor	-0.472	0.003
PV.Time	0.182	0.01

For the VAIs, preverbs of time also increased the likelihood of the Independent, but there are now variables that increase the chance of observing a Conjunct form. Discourse preverbs were those that most strongly increased the chance of a Conjunct, closely followed by obviative actors. Less strongly affecting a Conjunct form is the sg.actor effect. This image of the Conjunct as a form dealing with a preverb of discourse and a non-proximal actor suggests that the Conjunct is an Order that represents a construction beyond simple declarative clauses.

(17) êkwa, wîhkât nânitaw **kâ-isi-mâyinikêhkâto-cik** ôki nêhiyawak ...
and, ever simply CNJ-thus-act.badly.towards.each.other-3.PL these Cree ...

'And it was rare for the Crees to commit any crimes against one another at that time ...' (Vandall and Douquette, 1987, 47)

In (17) we see a Conjunct type verb, *kâ-isi-mâyinikêhkâtocik*, which takes the discourse preverb *isi-*. A large number of the Independent forms in this alternation and inflectional class are simply the quotative *itwêw*: 919 of 2157 tokens, to be exact (note that this ratio is much higher than in the Conjunct, where it is 209 of 4180 tokens). Despite this, the verb class AI.Speech was not found to be a significant effect on the Independent Order.

Transitive Inanimate Verbs

Table 5.26: Multivariate results for the Independent vs. Conjunct Alternation: VTIs

	Independent	
	Estimate	*p*-value
PV.Discourse	-2.064	< 0.001
TI.money.count	-1.914	0.02
NI.place.goal	-1.441	0.03
NDI.Body.goal	-1.032	0.04
actor.3	-0.793	< 0.001
NI.nominal.goal	-0.759	0.009
TI.do	-0.766	< 0.001
actor.2	0.372	0.04
NA.persons.actor	0.495	0.001
Px1Sg.goal	1.613	0.005

The VTIs had many significant effects. As with the other classes, the majority of these effects showed influence towards a Conjunct paradigm. Again the Conjunct Order is associated with with the discourse level. Verbs of action and verbs of money/counting also increased the likelihood of a Conjunct form. Place goals, nominalized goals, and body part goals similarly increased the likelihood of the Conjunct, as did third person actors. Person/human actors and especially second person actors, as well as those with goals possessed by first persons all increased the chance of observing an Independent form. This suggests the Independent to be an Order more related to local participants or those dependent on them as in (18), while the Conjunct is more likely to have an overt goal and a non-local actor, as in (19).

(18) ... **ki-kiskêyihtê-nâwâw** kîstawâw ...
... 2.PL.know.it-2.PL 2.PL.also ...

'... **you all know this** ...' (Ahenakew, 2000, 40)

(19) ... kayâs ayis ês ... **ê-kî-wêpina-hkik** ... wîwatiwâwa ...
... long ago for evidently ... CNJ-PST-throw.it.away-3.PL ... their medicine-bundles ...

'... for long ago evidently **they had thrown away** their medicine-bundles ...' (Ahenakew, 2000, 164)

Notably, unlike the VII and VAI classes, preverbs of time were not significant effects for either outcome. Curiously, the prescence of overt goals of any sort were not significant for the Independent Order. Perhaps the most striking aspect of these results, is that no semantic class of overt goals produced a significant effect in modelling the Independent. It is unclear why this might be, though the fact that the Conjunct outcome had more than double the number of observations than the Independent may have some impact.

Transitive Animate Verbs

Table 5.27: Multivariate results for the Independent vs. Conjunct Alternation: VTAs

	Independent	
	Estimate	*p*-value
TA.food	-1.723	0.007
PV.Position	-1.026	0.01
actor.obv	-0.921	< 0.001
PV.Move	-0.612	0.01
PV.Time	-0.342	< 0.001
Sg.actor	-0.608	0.04
goal.2	-0.487	0.03
NA.persons.goal	-0.352	0.01
goal.obv	-0.314	0.03
actor.1	0.485	< 0.001

For VTAs, verbs which regarded food strongly motivated Conjunct forms. position preverbs, movement, and time all increased the likelihood of the Conjunct Order, as did obviative actors/goals, person goals, and singular actors. Only one effect was associated with the Independent in the VTAs, that of first person actors. In this class, it seems the Conjunct Order is, unlike in other cases, non-present in nature, as well as being modified by preverbs. The independent still associated with a local actor, but not second person. The Conjunct is associated with obviative fits with the VII and VAI classes.

(20) ... wâposwa ê-kî-nipahât **ê-wî-kakwê-asam-iko-yâhk** wiya ... (C8GB 13)
... rabbits s/he kills him CNJ-FUT.VOL-try-feed-INV-3SG.1PL for ...

In (20)[6] we see a VTA, *ê-wî-kakwê-asam-ikoyâhk*, that represents not only a verb of food and eating, but also a preverb of time.

As noted, the *only* significant effect for the Independent VTAs was first person actors. This discrepency may simply be written off as an issue of data sparsity, as there were 1071 Independent TAs in this alternation and 1931 Conjunct forms, though this difference in data size is not so big that one would expect numerous effects to be significant for the Conjunct while only one was so for the Independent.

5.3.2 Independent vs. ê-Conjunct

As in the previous alternation, positive effects represent an influence of an effect on producing an Independent form; here, however, negative effects represent an increase in likelihood specifically of the ê-Conjunct form.

Intransitive Inanimate Verbs

Table 5.28: Multivariate results for the Independent vs. ê-Conjunct Alternation: VIIs

	Independent	
	Estimate	(*p*-value)
(Intercept)	-0.932	< 0.001
PV.Time	0.654	0.003

As in the alternation between the Independent and the general Conjunct, the VIIs had only one significant effect: preverbs of time, which increased the likelihood of observing an Independent form.

[6]The citation here of *C8GB 13* refers to the line number of the corpus file used to find this, as an exact page in a printed document was unavailable at time of writing. See the prefatory *Corpus Abbreviations* page in the front matter of this dissertation to determine what Corpus code is associated with which corpus file

Intransitive Animate Verbs

Table 5.29: Multivariate results for the Independent vs. ê-Conjunct Alternation: VAIs

	Independent	
	Estimate	(*p*-value)
(Intercept)	-1.529	< 0.001
PV.Discourse	-1.087	< 0.001
PV.Time	0.204	0.008
actor.3	0.373	0.007
actor.1	0.632	< 0.001

In the VAIs, discourse preverbs again strongly increased the likelihood of ê-Conjunct forms in the VAIs, while all other significant effects increased the likelihood of the Independent Order: preverbs of time and first or third person actors. These results suggest again an Independent form which is more focused in simple declarative structures that may be in a tense other than present, as in (21), where the main verb *kî-atoskêw* takes an Independent form.

(21) êwakw âna mâna nisis, Sam Minde, **kî-atoskê-w** pêyakwan
There it is that one habitually father-in-law's brother, Sam Minde, PST-work-3.SG similar

âta kâ-minihkwê-t
although CNJ-drink-3.SG.

'My father-in-law's brother, Sam Minde, **still used to work** the same, even when he drank.'

(Minde, 1997, 102)

There are no significant effects in the form of semantic classes of verbs. This is true even in spite of the fact that nearly 21% of all Independent VAIs were forms of *itwêw*, 'S/he says.' This quotative is tagged as a AI-Speech verb, yet this effect was not found to be significant in the logistic models for the Independent vs. ê-Conjunct alternation.

Transitive Inanimate Verbs

Table 5.30: Multivariate results for the Independent vs. ê-Conjunct Alternation: VTIs

	Independent	
	Estimate	(*p*-value)
PV.Discourse	-2.366	< 0.001
TI.money.count	-2.029	0.02
NI.place.goal	-1.604	0.02
TI.do	-0.912	< 0.001
actor.3	-0.797	< 0.001
NA.persons.actor	0.583	0.001
actor.2	0.851	< 0.001
Px1Sg.goal	1.275	0.04

For VTIs in the Independent vs. ê-Conjunct alternation, preverbs of discourse again strongly increased the likelihood of an ê-Conjunct form. Unlike the previous verb classes, VTIs also contained significant effects in terms of semantic classes. Verbs of money and action, as well as goals representing places, all increased the likelihood of an ê-Conjunct form. Additionally, third person actors corresponded to the ê-Conjunct outcome. Actors representing people, second person actors, and goals possessed by first person actors all increased the likelihood of the Independent Order.

(22) êkoni kahkiyaw ê-kî-wâpahta-mân tânis âya **ê-kî-isi-pamina-hkik** kîkway ...
those all CNJ-PST-see.it-1.SG how hm CNJ-PST-thus-look.after.it-3.PL something ...

'I saw all these things, how **they looked after** things ...' (Minde, 1997, 96).

In (22), *ê-kî-isi-paminahkik* represents a third person action verb with a discourse preverb that heads a non-main clause and occurs in the ê-Conjunct. It is worth noting, however, that the main verb in this excerpt, *ê-kî-wâpahtamân* is also in the ê-Conjunct. Using the information from Cook (2014) and the hypothesis that the Conjunct in general is less *main* clause-like or indexical than the Independent, one might expect that this token of *wâpahtam* to occur in the Independent.

Transitive Animate Verbs

Table 5.31: Multivariate results for the Independent vs. ê-Conjunct Alternation: VTAs

	Independent	
	Estimate	(*p*-value)
TA.food	-1.648	0.003
PV.Position	-1.342	0.001
PV.Discourse	-1.108	0.003
actor.obv	-0.801	0.01
PV.Move	-0.506	0.03
PV.Time	-0.331	0.001
actor.1	0.539	0.001
actor.2	1.807	< 0.001

In the Transitive Animate Verb class, verbs to do with food and preverbs of discourse strongly influenced a verb to occur in the ê-Conjunct. Preverbs of discourse and position, along with verbs with an obviative actor, had mild effects in influencing the ê-Conjunct. More mild effects in the form of preverbs of movement and time were also present. Local actors had moderate to strong effects, with second person actors having the most strong effect, increasing the likelihood of the Independent. This again suggests the ê-Conjunct as a marked form associated with discursively marked and less proximate actions, as well as those displaced in time. This is reflected in (23), where the main verb, *ê-kî-ayi-mâkohikot* is given in the ê-Conjunct.

(23) iyikohk mâna **ê-kî-ayi-mâkoh-iko-t** anihi wîhtikowa tâpwê ...
so much truly CNJ-PST-ah-is.pressed.upon-INV-3SG.3OBV that windigo.OBV truly ...

'And **he was truly pressed upon** by that windigo ...' (Ahenakew, 2000, 34)

5.3.3 Conjunct Type

The final alternation detailed in this section is multinomial: ê-Conjunct vs. kâ-Conjunct vs. Other-Conjunct forms. As a result, while a positive effect for an ê-Conjunct outcome does represent an increased likelihood of ê-Conjunct forms, a negative effect for the same

effect can not be interpreted as an effect toward some other specific outcome as in the previous alternations. Instead, a negative effect can simply be said to represent a decrease in likelihood for a given outcome. This is because, while in previous alternations there were only two possible options (and thus the absence of one implies the presence of the other), in multinomial results framed through a one-vs-rest heuristic, the absence of one outcome implies the presence of *any* other possible outcome. In the tables below, the estimates are given in each cell, with a *p* value being given underneath in parenthesis.

Intransitive Inanimate Verbs

Table 5.32: Multivariate results for the Conjunct Type Alternation: VIIs

	ê-Conjunct	**kâ-Conjunct**	**Other-Conjunct**
	Estimate (*p*-value)	Estimate (*p*-value)	Estimate (*p*-value)
(Intercept)	1.468 (0.004)	-2.554 (< 0.001)	-2.452 (< 0.001)
II.weather		1.596 (0.02)	

In Intransitive Inanimate Verbs, there was a single significant effect: weather verbs strongly increased the likelihood of the kâ-Conjunct. This effect was not significant for other outcomes.

(24) mâka mân ânohc **kâ-kîsikâ-k** kâ-mâmitonêyihtamân ...
but used to today CNJ-today-3.SG I think about it ...

'But when I think of it **today** ...' (Bear et al., 1992, 218)

(25) ... âta **kâ-kimiwa-hk** ...
... although CNJ-rains-3.SG ...

'... even when **it was raining** ...' (Minde, 1997, 36)

In (25), *kâ-kîsikâk* is used adverbally as an adjunct of time. The kâ-Conjunct here seems to represent a non-hypothetical conditional form, as opposed to the relativized form as in other instances.

Intransitive Animate Verbs

Table 5.33: Multivariate results for the Conjunct Type Alternation: VAIs

	ê-Conjunct	**kâ-Conjunct**	**Other-Conjunct**
	Estimate (*p*-value)	Estimate (*p*-value)	Estimate (*p*-value)
(Intercept)	0.923 (< 0.001)	-1.342 (< 0.001)	-3.052 (< 0.001)
actor.2	-1.147 (< 0.001)		1.849 (< 0.001)
Sg.actor	-0.695 (0.001)	0.633 (0.004)	
actor.1	0.471 (< 0.001)	-0.553 (< 0.001)	
NDA.Relations.actor	0.561 (0.03)		-2.744 (0.01)
actor.3	0.563 (< 0.001)	-0.653 (< 0.001)	
RdplW	0.616 (0.03)		
PV.Discourse	0.791 (0.003)	-0.719 (0.03)	
PV.Position	1.101 (0.001)	-0.985 (0.01)	
PV.WantCan		-1.227 (0.05)	1.83 (< 0.001)
AI.cooking			
PV.Qual			-1.637 (0.03)
AI.health			1.359 (0.005)
NA.persons.actor		0.379 (0.04)	
Pl.actor		0.578 (0.02)	

As previously, effects were far more numerous for VAIs than VIIs. Second person actors decreased the likelihood of the ê-Conjunct while increasing the likelihood of the Other-Conjunct class. Singular actors decreased the likelihood of the ê-Conjunct

but increased the likelihood of the kâ-Conjunct, as did third person actors, preverbs of discourse, and position preverbs (the latter most strongly). Actors belonging to the class of dependent relations increased the likelihood of ê-Conjunct but strongly decreased the likelihood of the Other-Conjunct class. The final class which affects the ê-Conjunct is the presence of weak/light reduplication, which acted as an effect for no other outcome. Preverbs of desire/ability strongly decreased the likelihood the kâ-Conjunct and similarly increased the likelihood of the Other-Conjunct. Verbs of cooking moderately decreased the likelihood of the kâ-Conjunct. Preverbs of quality and verbs of health had strong effects on the Other-Conjunct, the former a negative effect and the latter a positive. Finally, actors representing people and plural actors more generally had a moderate effect increasing the likelihood of a kâ-Conjunct.

(26) ... êkosi namôya ki-kiskêyihtê-nânaw tânitê **ê-isi-pimohtê-cik** êkwa kitôskâyiminawak ...
... êkosi NEG PST-know-21.PL where CNJ-thus-walk-3.PL and our kids ...

'... so we do not know where **our young people are going** ...' (Vandall and Douquette, 1987, 42-43)

(27) â, anohc kâ-kîsikâ-k, êwak ôhc êtikwê ayisiyiniw **kâ-maskawâtisi-t** ...
â, today CNJ-day-3.SG, this from perhaps person CNJ-be.strong-3.SG ...

'Well, that must be the reason why people **are so strong** today ...' (Bear et al., 1992, 364)

(28) tânisi k-êtôtamân, **mêstohtê-yêko** pê-miyi-kawi-yâni wêpinâson ...
what I will do, die-2PL.CNJ.FUT.COND come-give-INV-1PL.3SG flag ...

'What will I do **when you are all gone** if someone comes and gives me cloth ...' (Bear et al., 1992, 128)

In this alternation as in others, the ê-Conjunct is associated with first and second persons, and those with preverbs of discourse and position, as in (26). The majority of the kâ-Conjunct effects were negatively related, with the only positive effects being actor based: Sg.actor, Pl.actor, and NA.persons.actor, as seen in (27). The effects of the Other-Conjunct do not seem to form a cohesive class, though a verb with a second person actor and verb of health (in this case, *mêstohtêyêko*, meaning 'when you have all passed away'), is represented in (28).

Transitive Inanimate Verbs

Table 5.34: Multivariate results for the Conjunct Type Alternation: VTIs

	ê-Conjunct Estimate (*p*-value)	**kâ-Conjunct** Estimate (*p*-value)	**Other-Conjunct** Estimate (*p*-value)
(Intercept)	1.468 (< 0.001)	-2.361 (< 0.001)	-2.652 (< 0.001)
actor.2	-1.21 (< 0.001)		1.847 (< 0.001)
Prox.goal	-1.039 (< 0.001)	0.898 (0.001)	
PV.WantCan	-1.031 (0.003)	1.365 (< 0.001)	
TI.speech	-0.776 (0.02)	0.817 (0.02)	
NI.nominal.goal	-0.753 (0.007)		0.86 (0.01)
Sg.goal	-0.479 (0.02)		
PV.Position	2.362 (0.002)	-2.19 (0.04)	-2.203 (0.05)
NA.persons.actor		0.791 (0.002)	
NI.object.goal			-0.988 (0.004)
Med.goal			0.983 (0.01)

In the Conjunct Type alternation for VTIs, second person actors strongly decreased the likelihood of the ê-Conjunct while similarly increasing the likelihood of an Other-Conjunct outcome. Proximate goals, preverbs of desire and ability, and verbs of speech all strongly decreased the likelihood of the ê-Conjunct while strongly increasing that of the kâ-Conjunct. Nominalized goals strongly decreased the likelihood of the ê-Conjunct, but instead of being significant in the kâ-Conjunct outcome, this effect strongly increased the likelihood of the Other-Conjunct. Singular goals had a moderate negative effect on the ê-Conjunct outcome alone. position preverbs had extremely strong effects for all outcomes:

positive for the ê-Conjunct and negative for the other outcomes. Person actors strongly increased the likelihood of the kâ-Conjunct, and inanimate objects negatively influenced the Other-Conjunct outcome. Finally, medial goals strongly increased the likelihood of the Other-Conjunct class.

These results do not create clear profiles for these outcomes. What can be abstracted is that the ê-Conjunct is less likely to be used with proximal goals, less likely to be a verb of speech, and more likely to indicate something related 'position' (metaphorically extended to indicated negativity as previously described), as in (29); that the kâ-Conjunct is more likely to have a proximal goal, have a person actor and *not* have a position preverb (as in (30)); and the Other-Conjunct outcome has a second person actor, a medial goal, but *not* a position preverb (as in (31)).

(29) ... môy **ê-ohci-kaskihtâ-yâhk** ka-kîsowihkaso-yâhk ...
... no CNJ-PST.NEG-manage-1.PL CNJ-get.warm-1.PL ...

'... **we did not manage** to get warm ...' (Minde, 1997, 116)

(30) ôhi wiya kayâhtê ayisiyiniwak **kâ-kî-âpacihtâ-cik** ...
this that before people CNJ-PST-use-3.PL ...

'The things that **people used to use** formerly ...' (Bear et al., 1992, 294)

(31) kwayask êkwa anita **ta-kakwê-pimipayihtâ-yêk** anima kâ-nêhiyawê-yêk ...
effort and there CNJ-try-keep.up-2PL this CNJ-speak.cree-2.PL ...

'**you should make a serious effort** to keep speakings your Cree ...' (Whitecalf, 1993, 26)

Transitive Animate Verbs

Table 5.35: Multivariate results for the Conjunct Type Alternation: VTAs

	ê-Conjunct Estimate (*p*-value)	**kâ-Conjunct** Estimate (*p*-value)	**Other-Conjunct** Estimate (*p*-value)
(Intercept)			-2.601 (0.02)
actor.2	-1.595 (< 0.001)		1.829 (< 0.001)
Prox.actor	-0.938 (0.02)	1.324 (0.001)	
Sg.goal	-0.46 (0.04)	0.585 (0.01)	
actor.1	0.432 (0.007)		-0.751 (0.001)
Px1Sg.goal	0.695 (0.05)		
PV.Discourse	1.359 (< 0.001)		-2.463 (0.02)
PV.Position	1.775 (0.004)	-1.459 (0.05)	
TA.cognitive		-0.428 (0.05)	0.667 (0.02)

Results for the VTAs were similar to those for VTIs. Second person actors strongly decreased the likelihood of the ê-Conjunct and increased the likelihood of the Other-Conjunct outcome. Proximate actors and singular goals had strong and moderate effects (respectively) decreasing the likelihood for the ê-Conjunct outcome and increasing the likelihood of the kâ-Conjunct. First person actors mildly increased the likelihood of the ê-Conjunct and had a strong negative effect for the Other-Conjunct outcome. Goals possessed by singular first persons had a strong effect for the ê-Conjunct. Preverbs of discourse had a strong positive effect for the ê-Conjunct and a very strong negative effect for the Other-Conjunct. Position preverbs strongly increased the likelihood of the ê-Conjunct and decreased the likelihood of the kâ-Conjunct. Finally, the only verbal semantic class that showed a significant effect were those verbs of cognition, which had

a moderate negative effect on the kâ-Conjunct and a strong positive effect on the Other-Conjunct.

This creates a profile wherein the ê-Conjunct is associated with first person actors as well as preverbs of discourse, similarly to the way the outcome is framed in the Independent vs. ê-Conjunct alternation. This is exemplified in (32). The kâ-Conjunct class had fewer positive effects, but a verb with a proximate actor (and lacking a position preverb) that has not to do with speech is presented in (33). Finally, the Other-Conjunct class as embodied by second person actors on verbs of cognition and a lack of discourse preverb is presented in (34).

(32) ... mêkosi piko **ê-isi-wîhtamâ-t-akok** ...
... that is all only CNJ-thus-tell-INV-1SG.2PL ...

'... that is all I am telling you ...' (Kâ-Nîpitêhtêw, 1998, 66)

(33) êwakw ânima kêhcinâ aya ê-kî-miywêyihta-mân, ê-kî-oh- aya **ê-kî-isi-wâpam-ak**
this that certainly well CNJ-PST-be.glad-1.SG ê-kî-oh- well CNJ-PST-thus-see-1.SG.3.SG

niwîkimâkan ôtê kâ-pê-wîcêwak ... (EM 65; 0.93)
my husband over here CNJ-come-1.SG.marry.3.SG

'I certainly used to be happy that I could see my husband in this light when I came over here to be married to him ...' (Minde, 1997, 36-37)

(34) ... **ka-kitâpamâ-yêkok** iskwêwak ôtê ê-sâkaskinêkâpawi-cik ...
... CNJ-look.at-2PL.3PL women over there CNJ-stand.crowded-3PL ...

'... for you watch these women standing crowded over there ...' (Kâ-Nîpitêhtêw, 1998, 126)

5.4 Model Statistics

In assessing the results detailed above, we must also scrutinize the predictive models that produce such results. Arppe (2013) provides a function for this, `modelstats`. This function reports detailed specifics of how models operate, how often they predict a correct outcome, measures of classification precision, recall, tau (τ) measures of how much better than the baseline proportions the model's classifications are, and pseudo-R^2 (ρ^2) measures, a measure of *reduction* is badness-of-fit.

In the following subsections, tables for each of the verb classes are given for each alternation. For the final, multinomial, alternation being studied, a table is given for each of the outcomes (e.g. ê-Conjunct vs. all other Conjunct forms (*other*)). In the model statistics tables, column names with hats over their entirety (e.g. $\widehat{\text{CNJ}}$ for the Conjunct and $\widehat{\text{IND}}$ Independent) represent how many times the model *predicted* an outcome. Rows without hatted titles (e.g. CNJ) represent observed outcomes used to train the model. The cells represent how often a token with a predicted outcome was actually observed with a particular outcome (e.g. the cell $C_{\widehat{\text{IND}}|\text{CNJ}}$ represents how many Conjunct tokens were predicted to be Independent.).

5.4.1 Independent vs. Conjunct

Table 5.36: VII Independent vs. Conjunct

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	562	44
IND	141	46
Accuracy	0.77	
τ	0.35	
ρ^2	0.13	
	CNJ	IND
Recall	0.93	0.25
Precision	0.80	0.51

Table 5.37: VAI Independent vs. Conjunct

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	3973	282
IND	1045	1037
Accuracy	0.79	
τ	0.53	
ρ^2	0.27	
	CNJ	IND
Recall	0.93	0.50
Precision	0.79	0.79

Table 5.38: VTI Independent vs. Conjunct

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	1982	176
IND	561	336
Accuracy	0.76	
τ	0.42	
ρ^2	0.16	
	CNJ	IND
Recall	0.92	0.38
Precision	0.78	0.66

Table 5.39: VTA Independent vs. Conjunct

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	1765	207
IND	541	494
Accuracy	0.75	
τ	0.45	
ρ^2	0.21	
	CNJ	IND
Recall	0.90	0.48
Precision	0.77	0.71

The models for the Independent vs. Conjunct alternation performed reasonably well. The VII model was reasonably accurate at 77%. The model had a 93% recall for Conjunct and a 25% recall for the Independent. Precision for each outcome are similarly disparate, at 80% for the Conjunct and 51% for the Independent. While the recall and precision scores for the Conjunct outcome seem to suggest an accurate model, the Independent rates suggest a more mediocre model. The ρ^2 measure of 0.12 similarly suggests a middling model, as did a τ measure of 0.35 (suggesting model is somewhat better than baseline classification, not not substantially so). The VAI model showed similar accuracy (79%),

Conjunct recall (93%), and a Conjunct precision (79%), though Independent recall was higher at 50% and 79%. The τ and ρ^2 measures were also notably higher: the former at 0.53 and the latter at 0.27. These measures suggest a model with a decent increase over baseline in terms of classification and a large reduction in badness of fit (thus reflecting a model that well describes the variation). The VTI model showed a similar profile as the VII, with an overall accuracy of 76%, a Conjunct recall of 92% and precesion of 78%, with an Independent recall of 38% and precision of 66%. The VTI τ of 0.42 suggest a somewhat mediocre increase in classification over baseline, while the ρ^2 of 0.16 suggest a mediocre model fit. Conversely, the VTA model patterns more closely to the VAI model. The VTA model had an overall accuracy of 75%, a Conjunct recall of 90%, a Conjunct precision of 77%, an Independent recall of 48% and an Independent precision of 0.71. While the τ is mediocre at 0.45, the ρ^2 of 0.21 represents a relatively well fit model.

5.4.2 Independent vs. ê-Conjunct

Table 5.40: VII Independent vs. ê-Conjunct

	ê-CNJ	IND
ê-CNJ	378	33
IND	131	73
Accuracy	0.73	
τ	0.40	
ρ^2	0.17	
	ê-CNJ	IND
Recall	0.92	0.36
Precision	0.74	0.69

Table 5.41: VAI Independent vs. ê-Conjunct

	ê-CNJ	IND
ê-CNJ	2834	266
IND	987	1170
Accuracy	0.76	
τ	0.51	
ρ^2	0.27	
	ê-CNJ	IND
Recall	0.91	0.54
Precision	0.74	0.82

Table 5.42: VTI Independent vs. ê-Conjunct

	ê-CNJ	IND
ê-CNJ	1322	193
IND	468	490
Accuracy	0.72	
τ	0.44	
ρ^2	0.20	
	ê-CNJ	IND
Recall	0.85	0.55
Precision	0.74	0.73

Table 5.43: VTA Independent vs. ê-Conjunct

	ê-CNJ	IND
ê-CNJ	1083	269
IND	409	662
Accuracy	0.72	
τ	0.43	
ρ^2	0.22	
	ê-CNJ	IND
Recall	0.80	0.62
Precision	0.73	0.71

The models in the Independent vs. ê-Conjunct alternation were generally well fitting, with the exception of the VII model. This model had an accuracy of 73%, an ê-Conjunct recall and precison of 92% and 74% respectively, and an Independent recall and precision of 36% and 69%. The model's τ measure was mediocre at 0.40 and its ρ^2 was similar at 0.17. The VAIs showed an increase in nearly all measures: accuracy was 76%, ê-Conjunct recall and precision were 91% and 74%, Independent recall and precision were 54% and 82%, τ was measured at 0.51 (representing a decent increase over baseline in classification), and its ρ^2 was a relatively high 0.27, representing a large reduction in

badness of fit. The VTI model was slightly less effective, with a 72% accuracy, an ê-Conjunct recall and precision of 85% and 74%, an Independent recall and precision of 55% and 73%. The model's τ score was mediocre at 0.44 and a relatively robust ρ^2 of 0.20, representing a good fit. The VTA model's accuracy was 72%, its ê-Conjunct recall and precision were 80% and 73%, and Independent recall and pecision were 62% and 71%. Finally, the VTA model had a τ 0.43, representing a slight increase in classification over baseline, and a ρ^2 of 0.22, representing a very good reduction in badness of fit.

5.4.3 Conjunct Type

Table 5.44: VII Conjunct Types

	ê–	kâ–	other–
ê-CNJ	375	36	0.00
kâ-CNJ	59	101	1
other	16	5	3
Accuracy	0.80		
τ	0.56		
ρ^2	0.34		
	ê–	kâ–	other–
Recall	0.91	0.63	0.13
Precision	0.83	0.71	0.75

Table 5.45: VAI Conjunct Types

	ê–	kâ–	other–
ê-CNJ	3062	31	7
kâ-CNJ	733	76	10
other	254	4	55
Accuracy	0.76		
τ	0.42		
ρ^2	0.21		
	ê–	kâ–	other–
Recall	0.99	0.09	0.18
Precision	0.76	0.69	0.76

Table 5.46: VTI Conjunct Types

	ê–	kâ–	other–
ê-CNJ	1475	24	16
kâ-CNJ	261	46	13
other	218	9	50
Accuracy	0.74		
τ	0.43		
ρ^2	0.25		
	ê–	kâ–	other–
Recall	0.98	0.14	0.18
Precision	0.76	0.58	0.63

Table 5.47: VTA Conjunct Types

	ê–	kâ–	other–
ê-CNJ	1317	24	11
kâ-CNJ	326	63	7
other	151	20	43
Accuracy	0.73		
τ	0.42		
ρ^2	0.20		
	ê–	kâ–	other–
Recall	0.97	0.16	0.20
Precision	0.73	0.59	0.71

The Conjunct Type models performed reasonably well. In all inflectional classes, the ρ^2 value indicated a significant amount of variation (with the VIIs reporting the highest score at 0.34, followed by the VTIs at 0.25, then the VAIs at 0.20, and finally the VTAs with a ρ^2 value of 0.20). Models were similarly ordered with regards to their τ values being 0.56 for the VIIs, 0.43 for the VTIs, 0.42 for the VAIs, and 0.42 for the VTAs. Accuracy deviated from this ordering. The VII model had an accuracy of 80%, the VAIs had an accuracy of 76%, the VTIs had an accuracy of 74%, and the VTAs had an accuracy of 73%. In general, precisions were reasonably high for the ê-Conjunct outcome, with slightly lower rates for the other two outcomes: for the VII, precision was 83% in the ê-Conjcunt, 71% in the kâ-Conjunct, and 75% in the Other-Conjunct; for the VAIs there were precision scores of 76%, 69%, and 76% for the ê-Conjunct, kâ-Conjunct, and Other-Conjunct respectively; for the same with VTIs there were precision scores of 76%, 58%, 63%; finally, for the VTAs there were precision scores of 73%, 59%, 71% for the ê-Conjunct, kâ-Conjunct, and Other-Conjunct respectively. Overall, recall was significantly less robust than precision, except in the ê-Conjunct. Worthy of note, the kâ-Conjunct had substantially lower scores for precision and recall for all classes. This, combined with the extremely high ê-Conjunct recall scores, suggests the models were not confident in predicting a non-ê-Conjunct form. Despite this, the ρ^2 indicate that a substantial amount of variation is explained by the predictors used in the models. Though somewhat contradictory when considering the recall and precision scores, it is possible that these high ρ^2 values are primarily the result of the fact that the ê-Conjunct is simply substantially more frequent. Given this, any model would classify successfully if it predicted an ê-Conjunct by default, regardless of the predictors used. That variation can be adequately explained primarily by frequency suggests that the Conjunct Type alternation may not be motivated by morphosyntactic predictors for the most part.

5.4.4 General Fitting

In addition to the actual results of alternation modelling, one can assess the performance of a model as compared to other possible models, as in Arppe (2008, 2009). By creating different models, each containing certain subsets of effects, one is able to use the ρ^2 and τ scores (as discussed above) to determine relative important of different effect subsets in explaining an alternation. If, for example, a model with only semantic effects has nearly the same ρ^2 and τ values as a model with both semantic and morphological effects, one can deduce that morphological effects provide little additional value in explaining an alternation. For the purposes of this dissertation, this technique allows us to investigate the efficacy and necessity of the mixed-effect models previously discussed, as opposed to simpler statistical models.

To this end, this section will compare five different models: those which **did not** include the random effect `Lemma` (dubbed SEM.MORPH because they included only fixed effects in the form of semantic and morphological information), those which only included the random effect `Lemma` (abbreviated LEM), those which included semantic fixed effects alongside the random effect of `Lemma` (abbreviated SEM.LEM), those which had the random effect of `Lemma` but only *morphological* effects as fixed effects (abbreviated MORPH.LEM), and finally those full mixed-effect models (abbreviated ME) which include both types of fixed effects and the random effect.[7] Table 5.48 depicts the composition of the models to be compared.

[7]Because the relative performance of semantic/morphological information is apparent when comparing SEM.LEM and MORPH.LEM models, and because it is already clear the extent to which lemma identity is paramount to model performance, models featuring only semantic effects or only morphological effects (without the random effect of lemma identity) are not included in this comparison

Table 5.48: Model Composition

	Lemma (random)	Morphological (fixed)	Semantic (fixed)
SEM.MORPH		X	X
LEM	X		
SEM.LEM	X		X
MORPH.LEM	X	X	
ME	X	X	X

For this dissertation, morphological effects are those which have obvious and easily identifiable morphological exponents, such as `RdplS` or `goal.1`. Although some effects are specified for `actor` or `goal`, these tags are not considered semantic as they are relatively clearly associated with a suffix or suffix chunk. Below is a list of all morphological effect used in any model:

```
actor.1
actor.2
actor.3
actor.4
D.goal
goal.1
goal.2
goal.3
goal.obv
Pl.actor
Pl.goal
Px1Sg.goal
Px3Pl.goal
Px3Sg.goal
RdplW
Sg.actor
Sg.goal
```

Semantic effects are defined as those which do not have clear morphological exponents. This includes semantic classes, preverb groups, and descriptions of arguments

(e.g. `dem.goal` for goals which are demonstrative). Below is a list of semantic effects used throughout modelling.

```
AI.cooking
AI.do
AI.health
AI.pray
AI.reflexive
AI.speech
AI.state
Dem.actor
Der.Dim.goal
II.natural.land
II.sense
II.weather
Med.actor
Med.goal
NA.beast.of.burden.actor
NA.beast.of.burden.actor
NA.food.actor
NA.persons.actor
NA.persons.goal
NDA.Relations.actor
NDA.Relations.actor
NDI.Body.goal
NI.natural.force.goal
NI.nominal.goal
NI.object.actor
NI.object.actor
NI.object.goal
NI.place.goal
PV.Discourse
PV.Move
```

```
PV.Position
PV.Qual
PV.StartFin
PV.Time
PV.WantCan
PV.e
PV.kaa
PV.kaa
Prox.actor
Prox.goal
Prox.goal
TA.cognitive
TA.do
TA.food
TA.money.count
TA.speech
TI.do
TI.money.count
TI.speech
```

Table 5.49 details the difference in τ and ρ^2 between the five different types of models previously described for the Independent vs. Conjunct alternation.

Table 5.49: Model Comparisons. Independent vs. Conjunct, bold items represent a very good model fit, per McFadden (1973)

	VII		VAI		VTI		VTA	
	τ	ρ^2	τ	ρ^2	τ	ρ^2	τ	ρ^2
SEM.MORPH	0.31	0.05	0.47	0.15	0.32	0.09	0.35	0.10
LEM	0.36	0.12	0.52	**0.26**	0.34	0.15	0.42	0.18
SEM.LEM	0.35	0.12	0.52	**0.27**	0.36	0.15	0.44	0.19
MORPH.LEM			0.52	**0.27**	0.39	0.16	0.43	0.19
ME	0.35	0.12	0.53	**0.27**	0.42	0.16	0.45	**0.21**

In this alternation, it was nearly always the case that ME models had superior

performance in both classification improvement (τ) as well as reduction in badness of fit (ρ^2). This is not the case in two instances: the first is in the VII class, where LEM models appeared to have slightly higher τ and ρ^2 than the ME model, despite containing less information in terms of predictors. It is worth noting that the VII were the least numerous class, and due to its inherent semantics, it is significantly different than the other classes (in that it can refer to things like days of the week or temporal states). Also worth mentioning is that MORPH.LEM models were not available for the VII class, as the significant fixed effects for VIIs were all semantic. The other case is in the VTIs where the ρ^2 was the same for ME models as the MORPH.LEM models. In all cases, SEM.MORPH models showed significantly lower measures the other models, with LEM models showing a significant increase in ρ^2 and τ measures over SEM.MORPH models. This indicates that a significant amount of alternation is explained by random effects only. Put another way, individual lemmas appear to show significant variation in their propensity to occur in the Independent or Conjunct order more so than the use of fixed effects alone. The SEM.LEM and MORPH.LEM models create a slight increase in ρ^2 and τ over the LEM models,. While SEM.LEM and MORPH.LEM vary in which produces a better model depending on the verb inflectional class, this difference is usually minimal. Interestingly, all models except SEM.MORPH performed extremely similarly in the VIIs and VAIs, indicating that the influence of a random effect (Lemma) is one of, if not the, most important factor in modelling this alternation in these inflectional classes. As mentioned before, the ME (unsurprisingly) often perform better than any other model, and, excluding the VIIs, they are never *worse* than the other models in their ability to classify or reduce badness of fit.

For the Independent vs. ê-Conjunct alternation, a similar pattern to the previous alternation can be seen. In general, SEM.MORPH models provided very little explanation for the variation, while LEM models showed a significantly higher ρ^2. This again suggests that lemmas have inherent propensities to surface in one order over another. In all cases, ME models were the best fitting models, except for the VAIs where it was equal to the

Table 5.50: Model Comparisons. Independent vs. ê-Conjunct, bold items represent a very good model fit, per McFadden (1973)

	VII		VAI		VTI		VTA	
	τ	ρ^2	τ	ρ^2	τ	ρ^2	τ	ρ^2
SEM.MORPH	0.28	0.04	0.43	0.15	0.33	0.10	0.32	0.12
LEM	0.36	0.14	0.49	**0.26**	0.36	0.17	0.38	0.17
SEM.LEM	0.40	0.17	0.50	**0.27**	0.39	0.17	0.41	0.19
MORPH.LEM			0.50	**0.26**	0.41	0.19	0.44	**0.21**
ME	0.40	0.17	0.51	**0.27**	0.44	**0.20**	0.43	**0.22**

MORPH.LEM model. Similarly, all classes other than the VII showed had ME models with an $\rho^2 \geqslant 0.20$. The lower measure for the VII model is again likely the result of a paucity of data. Taken together, the model performance for the Independent vs. ê-Conjunct alternation suggest a generally better modeled alternation than in the Independent vs. Conjunct general alternation. Although theoretically one may conceive of Order as a class that is split principally between the Independent and Conjunct (as in Chapter 2), these results suggest that such an alternation is harder to model in terms of morpho-semantic properties. Instead, a more clear choice exists in whether one wants to use an Independent form or an ê-Conjunct form. This behaviour is not entirely unexpected, given the semantic differences of the different Conjunct types. Even Cook (2014) describes the ê-Conjunct as more of an elsewhere case and one more closely similar to the Independent in morphosyntactic behaviour. In fact, Cook (2014, 125) describes all conjunct type other then the ê-Conjunct as being disallowed from matrix clauses, which are the domain of the Independent and the ê-Conjunct. This comports with the performance of the above models. Further, the other Conjuncts are highly semantically representative in a way that does not need accounting for (and was not accounted for) in the models, as will be discussed below. Thus, mixing the ê-Conjunct together with these other, more straightforwardly described, forms as in the first alternation, may produce an outcome that a logistic model is simply not fully able to reproduce and explain.

Table 5.51: Model Comparisons. Conjunct Type Alternation, bold items represent a very good model fit, per McFadden (1973)

	VII		VAI		VTI		VTA	
	τ	ρ^2	τ	ρ^2	τ	ρ^2	τ	ρ^2
SEM.MORPH	0.33	0.11	0.37	0.05	0.38	0.07	0.36	0.07
LEM	0.55	**0.35**	0.37	0.18	0.40	**0.20**	0.35	0.16
SEM.LEM	0.56	**0.34**	0.38	0.19	0.41	**0.23**	0.38	0.17
MORPH.LEM			0.37	0.18	0.40	**0.20**	0.35	0.16
ME	0.56	**0.34**	0.42	**0.21**	0.43	**0.24**	0.42	**0.20**

In Conjunct type alternation, as depicted in Table 5.51, models performed variously. All ME models showed significant explained variation and improved classification (with $\rho^2 \geqslant 0.20$ and τ values of greater than 0.40). VAIs and VTAs were only successful in the ME case. Conversely, VIIs and VTIs were successful in all cases except for SEM.MORPH. These values indicate one requires semantic and morphosyntactic information along with lemma specific identities to accurately describe this Conjunct Type alternation. Specifically, it appears for the VIIs and VTIs, lemma information is paramount to accurate classification.

5.5 Exemplar Extraction

The fitted values of a mixed effects logistic model present a set of probability estimates for every outcome occurring in a particular context. For example, given the lemma token *itêw* for the VTA Independent vs. ê-Conjunct model in some explanatory context, the model estimates for the Independent construction type a probability of 0.85. Thus, there is an 85% chance that the verb will occur in the Independent Order; conversely, there is a 15% chance of it occurring in the ê-Conjunct. These probability estimates allow for the easy extraction of explanatory contexts most likely to be in one Order over another, which may prove useful in educating students of the language on how and when to use each Order.

Based on the work of Arppe (2008, 228-236), I first created dataframes for each inflectional class in each of the three alternations that are studied in this dissertation. These data frames include only the significant effects identified in Chapter 5. Hierarchical Agglomerative Clustering is used to create prototypical classes from which exemplars can be extracted. In contrast to an educated guess, but arbitrarily set, number of clusters used by Arppe (2008), silhouette analysis (Rousseeuw, 1987) was used to determine the appropriate number of clusters to be used. In this technique, a *silhouette* is calculated which represents the distance of a cluster member to other members of the same cluster, the distance of a cluster member to all objects in other clusters, and finally the distance of a cluster members to all members of the next nearest cluster (Rousseeuw, 1987). Using these measures, a silhouette is calculated which ranges from -1 (representing an item that certainly was misclustered) to +1 (representing an item that was certainly properly clustered). When the silhouette is equal to 0, it is unclear where the item belongs, and so can be said to lie between clusters (Rousseeuw, 1987).

Using the `fviz_silhouette` function from the `factoextra` library (Kassambara and Mundt, 2020) in `R`, I selected the number of clusters that created an average silhouette (the average silhouette for all classes divided by the number of classes) as close to 1.00 as possible. This results in some individual classes with low silhouettes, though it ensures the overall clustering is as well fit as possible. In some cases, an average silhouette of 1.00 could not be achieved even after a large amount of clusters were added. In this case, the highest possible silhouette (in all such cases this was a silhouette of 0.99) with the lowest number of clusters was chosen. In most cases, an optimal number of clusters ranged from 5-50, though in some cases such as the Independent vs. Conjunct VTAs, this number was much higher. Table 5.52 details the optimal number of clusters for each alternation and each class.

From these results, only those exemplars where the predicted outcome matched the *actual* outcome were considered. Only the five exemplars (or fewer, if there were

Table 5.52: Number of silhouettes used for clustering

		VII	VAI	VTI	VTA
IND V. CNJ		4	17	77	110
IND V. ê-CNJ		4	20	42	55
Cnj Type		4	20	42	55
	ê-CNJ	2	50	69	31
	Kâ-CNJ	2	53	22	16
	Other CNJ	2	13	21	11

fewer than five correctly predicted exemplars) with the highest probability estimates from each unique cluster (per *outcome*) are given. Note that, for the binary alternations, probabilities closer to 0 represent an alternative outcome (i.e. *not Independent*), while those closer to 1.00 represent the predicted variable outcome (i.e. *Independent*). For the sake of presentation, the alternative outcome probabilities in the binary cases will be given in the form of $1 - \texttt{probability}$, so that an estimate of 0.01 in the Independent vs. Conjunct alternation (indicating a likely Conjunct form) will be presented as a Conjunct probability of 0.99.

The following sections detail these exemplars. As before, sources are given by the corpus codes (e.g. *AL* for the corpus AL-RL-C.FIN) along with the line number where the exemplar occurs. A list of corpus codes and the full name of the corpus file is given in the prefatory material to this dissertation. Where translations are quoted verbatim from original published sources, the relevant book and page numbers are included. Next to the corpus ID and line number is the estimated probability. Where original translations were not available, only word-by-word glosses are given. The verbs which are being evaluated are in bold face. Exemplar verbs are fully morphologically glossed.

5.5.1 Independent vs. Conjunct

Inanimate Intransitive Verbs

Independent

(35) otâkosihk ma cî wiya **kî-pêhtâkwan** kwayask. (CMBK-5-2 20; 0.69)
yesterday NEG Q for PST-is.heard properly

(36) aya mîna **ni-kî-âtotê-n** kayâs, (AL 976; 0.54)
ah, and 1SG-PST-work-1SG long ago,

namôy wîhkâc **ohci-pêhtâkwan** ... (AL 976; 0.54)
not ever **it was heard** ...

There were only two classes containing VIIs with probability estimates of over 0.50 (that is, where the model predicted an Independent form). In both cases, these were past tense forms of *pêhtâkwan*, 'it is heard.' In the first instance, example (35), the verb is used in an interrogative clause.

Conjunct

(37) "tâpwê anim âkosi sâsay **ê-ispayi-k** anima kâ-kî-itwêt," itwêw ... (C2GB 40; 0.01)
truly that thus already CNJ-fares.thus-3.SG that he said, say-3.SG

' "It is true, and some of what he had said is happening already," she said ..."' (Bear et al., 1992, 80-81)

(38) êkosi anima mîna êwako **ê-kî-ispayi-k** mâna ... (SW 41; 0.97)
so that and that CNJ-PST-fares.thus-3.SG used to ...

'That is the way this used to happen ...' (Whitecalf, 1993, 36-37)

For the Conjunct, there were also only two classes with probability estimates of less than 0.50. Both instances are forms of the lemma *ispayin* 'it happens thus.' In both cases, the predicted verbs seem to carry the main semantic meaning of the utterance, and both appear to be in past tense forms, though in (37) this does not appear to be marked morphologically. This comports with the finding that the Conjunct Order is some how less immediate, as discussed previously. The Conjunct forms above are well predicted, with the lowest probability being 0.97.

Animate Intransitive Verbs

Independent

(39) vamps aniki, âha, 'asêsinwa' **kî-itwê-wak**. (AL 1144; 0.84)
vamps those yes 'asêsinwa' PST-say-3.PL

(40) ... nititwân mâna, tâspwâw mâna wiya niya **nit-itwâ-n** ... (SW 140; 0.82)
... I say usually, in fact usually by contrast I 1-say-1.SG ...

'... I usually say, as for myself, as a matter of fact, I usually say ...' (Whitecalf, 1993, 76)

(41) êkwa êkosi **kî-itwê-w** ana kisêyiniw ... (VDC2 1061-1062; 0.79)
and this PST-say-3.SG that old man ...

'and this is what that old man said ...' (Vandall and Douquette, 1987, 106-107)

(42) "a play ôm ê-wî-ayâyâhk ôtê Sandy Lake," **itwê-w** ... (AA 33; 0.75)
"a play then we are going to have over here Sandy Lake," say-3.SG ...

' "that we are going to have a play over here at Sandy Lake," he said ...'

(43) **ê-kî-wîhkitisit** mâna, ban- bannock ê-kî-osîhât ... (C6IC 12; 0.5)
CNJ-PST-taste.good-3.SG used to, ban- bannock she made it ...

'The bannock used to taste good, and she used to make it ...' (Bear et al., 1992, 148)

Of the top Independent exemplars, only four were accurately predicted and all were forms of the verb *itêw*, 's/he says.' Used as a quotative in each of these cases, the expected probabilities ranged from 0.75 to 0.84, suggesting decent confidence in the prediction of an Independent outcome.

Conjunct

(44) ... êkot[a] êkwa ki-kâh-kî-wîcêwâw tânis **ê-isi-mawimoscikê-t**. (JK 8; 0.97)
... there and you would join him how CNJ-thus-pray-3SG

'... then you would be able to join him in his way of worship.' (Kâ-Nîpitêhtêw, 1998, 51)

(45) ... otôsk-âyima êkâ kwayask **ê-isi-wîcêhto-yit**. (JK 6; 0.97)
... young people not properly CNJ-thus-join.together-3.OBV

'... if their young people do not get along with one another.' (Kâ-Nîpitêhtêw, 1998, 14)

(46) ... ôma tânisi ê-ispayik, tânisi **ê-kî-isi-mawimoscikê-cik** (VDC2 1050-1051; 0.97)
... this how it is, how CNJ-

nêhiyawak kayâs ... (VDC2 1050-1051; 0.97)
Cree long ago ...

(47) ... kîtahtawê êsa mâna êkwa kî-môyêyihtamwak ayisiyiniwak, tânêhki ohci anihi
... suddenly apparently used to and they were aware of it people why from that

mistahi **kâ-pê-kito-yit** êkota ... (SW 140; 0.96)
much CNJ-come-call-3.OBV then ...

'... and then people would realize why he had come to hoot there ...' (Whitecalf, 1993, 39)

(48) môy mâka wîhkâc ohci-wîhtam-wak awiyiwa anihi **kâ-kî-itahkamikisi-yit**, êha.
not still ever NEG.PST-tell.about-3.PL someone that CNJ-PST-behave.thus-3.OBV, yes.

(CMBK-5-2 110; 0.95)

Unlike the Independent outcome, there appears to be more variety in regards as to what lemmas occurred in the top five Conjunct exemplar forms. Further, estimated probabilities were much higher, with the model predicting all five exemplars with above a 95% confidence. In all cases, these exemplars showed clearly an subordination as described by Cook (2014) and Wolfart (1973). For example, (44), contains the exemplar verb *ê-isi-mawimoscikêt*. Here, *ê-isi-mawimoscikêt* references the way someone prays and is subbordinate to the main verb of this clause, *kikâh-kî-wîcêwâw* 'you would be able to join him'. Similarly, in (47) the predicted verb *kâ-pê-kitoyit* is a part of the relative clause subordinate to the main verb, *kî-môyêyihtamwak* 'They were aware of it.'

Transitive Inanimate Verbs

Independent

(49) "nikêhcinâhon ôki iskwêsisak, nikotwâw ê-kimotam-awi-cik," nititêyiht- **nit-itêyiht-ê-n**
"I am sure these girls, anytime they steal it from me," I thi 1-think-THM-1.SG

ôma niminihkwâcikan. (C7MW 75; 0.80)
this my cup

' "I am sure one of these girls has stolen it from me," I thought with respect to my cup.' (Bear et al., 1992, 180).

(50) **ki-kiskêyiht-ê-n** kiya? (AL 558; 0.80)
2-understand-THM-2.SG you?

(51) «kiya **ki-kaskiht-â-n** ê-osîhtâ-yan,» nit-it-ik. (AA 76; 0.74)
you 2-be.able-THM-2.SG you make it she says to me

‘ “You, you have been able to make it,” she says to me.’ (Ahenakew, 2000, 68-69)

(52) nîsta **ni-wî-nipahi-cîhkêyiht-ê-n** (CMBK-3-2 182; 0.76)
I too 1-FUT.VOL-very-like-THM-1.SG

(53) «kiyâm tâpwêhta, môy **ki-ka-mihtât-ê-n**,» ê-kî-isit mâna. (CMBK-4-2 114; 0.74)
please truly, not 2-FUT.CON-regret-THM-2.SG s/he says to me used to.

Independent VTI exemplars mostly had to do with cognition/emotion verbs with the {-êyi-} morph. The exception to this is in (51), where a form of *kaskihtâw*, ‘s/he is able to’ occurs in the Independent. The estimated probabilities were lower than above, ranging from 0.73-0.80

Conjunct

(54) kahkiyaw kîkway ‘mînisa’ k-êsiyîhkâtêki, nanâtohk **ê-kî-isi-osîht-â-t** kîkway wiyâs,
all thing berry it is called, variety CNJ-PST-thus-make-THM-3.SG what meat,

ê-osîhtât îwahikana ê-môwât. (VDC2 315-317; 0.00)
s/he makes it pounded meat s/he eats him.

‘All these things that are called “berries”, they prepared them in various ways, they prepared the meat and ate pounded meat.’ (Vandall and Douquette, 1987, 56-57)

(55) â, êkosi pêyakwâw êkota ê-kî-otahot ayi, wiyê [sic] nawac ê-kî-kiskêyihtahk ê-isi- kîkway
ah, so once there she beat me well, for [sic] before s/he knew it ê-isi- what

ê-isi-osîht-â-t. (C7MW 86; 0.99)
CNJ-thus-make-THM-3.SG

‘well, and so in that she knew better than I how to make something, this once she did beat me.’ (Bear et al., 1992, 192-193)

(56) êkwa aya, aya, pêyakwâw ê-kiskisiyân iyikohk ê-kî-miyokihtâyâhk askipwâwa, êkosi mân
and uh, uh, once I recall when we grew well potatoes, thus used to

ê-kî-isi-tipah-am-âhk, mitâtahtomitanaw-maskimot ê-kî-ayâyâhk (EM 117; 0.01)
CNJ-PST-thus-measure-THM-1.PL, one hundred bags we had it

‘And I remember once, when we grew such a good crop of potatoes, that is how we measured them, we had one hundred bags.’ (Minde, 1997, 84-85)

(57) “kây, êkâya mâto! ê-nôhtêhkatêt ana wîst ôm ê- ê-wâpamiko- ê-wâpahtahk
“no, do not cry! s/he is hungry that one s/he too it is this ê- ê-wâpamiko- s/he sees it

ôma wiyâs **ê-nôhtê-mîci-t**,” (C4MF 68; 0.99)
actually meat CNJ-want-eat-3.SG

‘ “Do not cry! That one is hungry, too, and it sees this meat and wants to eat it,” ’ (Bear et al., 1992, 112-115)

(58) ... ê-wî-nanâskomot, matotisân ôma **kâ-wî-osîht-â-t**. (JK 42; 0.98)
... he gives him thanks, sweat lodge actually CNJ-FUT.VOL-make-THM-3.SG

‘... that he is about to give thanks, the one who is about to make a sweat lodge.’ (Kâ-Nîpitêhtêw, 2007, 84)

The Conjunct exemplars for VTIs were mostly made of forms of *osîhtâw*, ‘s/he makes it,’ as in examples (54), (55), and (58). These verbs occurred in both main and subordinate clauses. Examples (56), *ê-kî-isi-tipahamâhk* contains a discourse preverb, while in (57) the exemplar verb *ê-nôhtê-mîcit* is desiderative. The Conjuncts were very well predicted with the lowest estimated probability being 0.99.

Transitive Animate Verbs

Independent

(59) sapiko mân êkosi **nit-it-â-wak** nôsisimak, «kayâs ôma niyanân mistahi
actually used to thus 1-say-DIR-1SG.3PL my grandkids, long ago FOC we much

ê-kî-atoskêyâhk ... » (cmbk-4-v2 304; 0.79)
we worked ...

(60) ... ômisi mâna **ni-kî-it-â-wak** nitawâsimisak ... (EM 66; 0.73)
... thus used to 1-PST-say-DIR-1SG.3PL my children

... I used to tell my children as follows ... (Minde, 1997, 36)

(61) “îwahikanak niwî-osîhâwak,” **nit-it-â-wak** awâsisak. (AL 407; 0.73)
“pounded meats I’m going to make them,” 1-say-DIR-1SG.3PL children

‘ “I’m going to make pounded meat,” ’ I told my children. (Bear et al., 1992, 206-207)

(62) “nôsisim!” **nit-it-ik** ... (C2GB 14; 0.70)
my grandchildren 1-say-INV.1SG.3SG ...

‘ “Grandchild!” she said to me ...’ (Bear et al., 1992, 68-69)

(63) ... miton ês âwa nôcikwêsiw, “ayiwêpitân,” **itê-w** êsa okosisa ... (C4MF 23; 0.63)
... very this this old lady let’s rest say-3SG.3OBV this her son ...

‘... and the old lady aid to her son, “Let’s rest;” ...’ (Bear et al., 1992, 106-107)

The top five Independent exemplars for the VTAs were all forms of *itwêw*, ‘s/he says to him/her.’ Specifically, each of these tokens were quotatives reporting exact speech. Perhaps unsurprisingly, these tokens were in either first or third person, but not second. Given the nature of the corpus (speeches or conversations). In all but one instance, (60), these quotatives all had present tense morphology. Outside of these five exemplars, there was one instance of a non-*itwêw* form occurring in the Independent and being correctly identified as such by the model, *kinisitohtâtinâwâw*, ‘I understand you all’; however, in this instance, the predicted probability was only 0.57, representing a relatively uncertain prediction. Generally, estimate probabilities were lower for this group than previous ranging from 0.63 to 0.79.

Conjunct

(64) îh, êwako anima êsa kayâs êkosi **ê-kî-pê-isi-kakêskim-â-cik**
look, this the fact that apparently long ago so CNJ-PST-come-thus-counsel-DIR-3SG.3PL

otôsk-âyimiwâwa ... (SW 140; 0.98)
their young over there

‘Look, in this wise long ago did they use to counsel their young people ...’ (Whitecalf, 1993, 76-77)

(65) ... kita-wâpamikot, **ê-pê-minihkwât-â-yit**, itwêw. (VDC2 485-486; 0.97)
... looking at him CNJ-come-drink-DIR-3.OBV he says

‘... looking at him, he said, to trade it for a drink.’ (Vandall and Douquette, 1987, 68-69)

(66) ... wâposwa ê-kî-nipahât **ê-wî-kakwê-asam-iko-yâhk** wiya ... (C8GB 13; 0.97)
... rabbits he killed CNJ-FUT.VOL-try-feed-INV-1PL.3SG they ...

(67) ‘... even though it looked as if they were not subject to any formal law when they did do something wrong.’ (Vandall and Douquette, 1987, 50-51)

(68) êkosi ôma aspin, “ay, kayâs nôcokwêsiw **ka-wayawî-pakamah-osk**!” nititikwak.
so that finally “hey, long ago old lady CNJ-outside-throw-INV.3SG.2SG!” they say to me

(VDC2-RES 561-562)

So at the end they say to me “hey, for sure, then, the old lady would throw you out, and with a vegence!” they say to me. (Vandall and Douquette, 1987, 74-75)

The Conjunct exemplars for the Independent vs. Conjunct VTA set were mostly subordinate verbs, such as *ê-pê-minihkwâtâyit* ‘comes to trade a drink for it’ (in (65)) or *ê-wî-kakwê-asamikoyâhk*, ‘to try to feed all of us’ (in (66)). These results comport with the general descriptions of Order in the literature. Exemplars were all highly predicted, with the Conjunct exemplars never being lower than 0.96.

5.5.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Independent

(69) ... otâkosihk ma cî wiya **kî-pêhtâkwan** kwayask. (CMBK-5-2 20; 0.85)
... yesterday not Q for PST-be.heard.3SG properly

... Was it not heard properly yesterday? (Ahenakew, 2000, 109-110)

(70) aya mîna kî-âtotên kayâs, namôy wîhkâc **ohci-pêhtâkwan** ... (AL 976; 0.74)
ah, and **I told** long ago, not ever NEG.PST-be.heard.3SG ...

As in the previous alternation, there were fewer than five exemplars available for each outcome in the VII class. Here, there are only two exemplars for the Independent. Given this number, it is hard to draw conclusions, though it is worth noting that in both cases the exemplar verb was a negative form of *pêhtâkwan*, ‘it is heard’. The difference between the two exemplars in estimated probabilities ranged from 0.74 to 0.85.

Conjunct

(71) “tâpwê anim âkosi sâsay **ê-ispayik** anima kâ-kî-itwêt...” (C2GB 40; 0.99)
“truly that thus already CNJ-fare.thus.3SG that s/he said...”

‘ “It is true, and some of what he had said is happening already ...” ’ (Bear et al., 1992, 80-81)

(72) êkosi anima mîna êwako **ê-kî-ispayik** mâna ... (SW 41; 0.92)
so that and this CNJ-PST-fare.thus.3.SG used to ...

‘That is the way this used to happen ...’ (Whitecalf, 1993, 37-38)

(73) ê-pânisamihk anima kahkiyaw, nama kîkway **ê-ohci-wêpinikâtê-k**. (EM 97; 0.50)
someone cuts it that all nothing CNJ-NEG.PST-be.wasted-3.SG

‘It was cut, and nothing was wasted.’ (Minde, 1997, 120).

The ê-Conjunct outcome had three exemplars, two of which are of the verb *ispayin*, ‘it happens.’ In (72) and (73) the exemplars are in past tense. Like the Independent, the Conjunct had a large range in estimated probabilities, ranging from 0.50 to 0.99.

Animate Intransitive Verbs

Independent

(74) ‘... môy kîhtwâm êkwa nika-pakitinâw wîhkâc awâsis,’ **nikî-itwâ-n** ôma (CMBK-3-2 170; 0.90)
... no again and I will let go never children 1-PST-say-1SG this

‘ “I will not ever let the children go again,” I said this’ (Ahenakew, 2000, 44)

(75) nititwân mâna, tâspwâw mâna wiya niya **nit-itwâ-n** ... (SW 140; 0.88)
I say usually, in fact usually for I 1-say-1SG

‘I usually say, as for myself, as a matter of fact, **I usually say** ...’ (Whitecalf, 1993, 76)

(76) êkwa êkosi **kî-itwê-w** ana kisêyiniw ... (VDC2 493-494; 0.88)
and so PST-say-3SG that old man ...

‘and this is what that old man **said** ...’ (Vandall and Douquette, 1987, 106-107)

(77) “a play ôm ê-wî-ayâyâhk ôtê Sandy Lake,” **itwê-w** ... (AA 33; 0.75)
“a play then we are going to have over here Sandy Lake,” say-3.SG ...

‘ “that we are going to have a play over here at Sandy Lake,” **he said** ...’ (Ahenakew, 2000, 44)

(78) ‘â, mahti! pâmwayês miton ôtâkosik, nika-nitawi-minihkwahastimwân’, k-êtwêyan,
‘well, please! before quite it is evening, I will water the horses,’ you said,

ki-kî-itwâ-n ... (CMBK-3-2 488; 0.82)
2-PST-say-2SG ...

As with the previous alternation, the VAI Independent exemplars were all forms of *itwêw*, ‘s/he said.’ In all examples other than (75), the exemplar verbs were semantically past, though not always morphologically so. Given the nature of quotatives, this is perhaps unsurprising. All exemplars were relatively well predicted, with estimated probabilities ranging from 0.89 and 0.88.

Conjunct

(79) ... otôsk-âyima êkâ kwayask **ê-isi-wîcêhto-yit**. (JK 7; 0.97)
.‘.. their young people not right CNJ-thus-cooperate-3.OBV.’

... if their young people do not **get along with one another**. (Kâ-Nîpitêhtêw, 1998, 48-49)

(80) ... êkot[a] êkwa kikâh-kî-wîcêwâw tânis **ê-isi-mawimoscikê-t**. (JK-C4ARC.798 8; 0.97)
... then and you would be able to how CNJ-thus-pray-3.SG

‘... then you would be able to **join him in his way of worship** (Kâ-Nîpitêhtêw, 1998, 50).’

(81) êkos êtikwê piko **ê-kî-isi-ma-mêyiwiciskê-hk** ê-kî-isi-pasikôhk. (C8GB 18; 0.96)
so apparently only CNJ-PST-thus-RDPLW-be.dirty-UNSPEC someone gets up.

‘... one simply **got up dirty**, I guess ...’ (Bear et al., 1992, 210-211).

(82) ê-kakwêcimak ôma, tânis **ê-kî-pê-ay-isi-pimâcihiso-cik** ayisiyiniwak ... (AL 2; 0.95)
I am asking her this, how CNJ-PST-come-RDPLW-thus-live-3.PL people

‘I am asking her this: **how people lived** ...’ (Bear et al., 1992, 240-241).

(83) tânitê kiy ê-kî-kiskinahamâkawiyan cî, ka-isi-kakâyawisî-yan **ê-awâsisîwiyan**? (AL 359;
where you you went to school Q, you work hard CNJ-be.child-2.SG?

0.95)

The exemplars for the Conjunct VAI nearly all made use of the {isi-} preverb, a preverb classified as a discourse preverb, indicating the ways in which an action was done. The only exemplar without this preverb was in (83) the simple second person ê-Conjunct form, *ê-awâsisîwiyan*, 'you ware a child.'

Transitive Inanimate Verbs

Independent

(84) **ki-kiskêyiht-ê-n** kiya? (AL 558; 0.89)
2-know-THM-2.SG you?

(85) kiya **ki-kaskiht-â-n** ê-osîhtâyan (AA 76; 0.88)
you 2-manage-THM-2.SG you prepare it

(86) kiyâm tâpwêhta, môy **ki-ka-mihtât-ê-n** (CMBK-4-2 114; 0.87)
so agree, not 2-FUT.CON-regret-THM-2.SG

(87) ... môy mâka wîhkât niya wiya **nôh-cîhkêyiht-ê-n** ôma 'radio'
... not but ever I for 1.NEG.PST-like-THM-1.SG this 'radio'
k-êsiyîhkâtêk (CMBK-4-2 128; 0.87)
it is called thus

(88) ... wiya kiyânaw **ki-kaskiht-â-naw** kîkway ka-kî-nipahtamâsoyahk ... (C2GB 45; 0.80)
... FOC we all 2-succeed-THM-21.PL what we all killed it for ourselves

As previously, in examples (85), (86), and (87) demonstrate the use of the Independent as the main verb in a multiverb phrase. Interestingly, all examples but (87) were in second person, and all forms had speech act participants as actors. Estimated probabilities range from 0.87 and 0.90.

Conjunct

(89) kahkiyaw kîkway 'mînisa' k-êsiyîhkâtêki, nanâtohk **ê-kî-isi-osîht-ât** kîkway wiyâs,
all thing berry it is called, variety CNJ-PST-thus-make-3SG.3OBV what meat,
ê-osîhtât îwahikana ê-môwât. (VDC2 315-317; 0.00)
s/he makes it pounded meat s/he eats him.

'All these things that are called 'berries', **they prepared them** in various ways, they prepared the meat and ate pounded meat ...' (Vandall and Douquette, 1987, 56-57).

(90) êkwa aya, aya, pêyakwâw ê-kiskisiyân iyikohk ê-kî-miyokihtâyâhk (EM 117; 0.99)
and uh, uh, once I recall when we grew well ...

askipwâwa, êkosi mân **ê-kî-isi-tipaha-mâhk**, mitâtahtomitanaw-maskimot ê-kî-ayâyâhk
potatoes, thus used to CNJ-PST-thus-measure-1.PL, one hundred bags we had it

(EM 117; 0.01)

'And I remember once, when we grew such a good crop of potatoes, that is **how we measured them**, we had one hundred bags ...' (Minde, 1997, 84-85)

(91) ... mêtoni mân **ê-kî-kanâcihtâ-cik** êkwa mân ê-kî-kaskâpasahkik. (EM 268; 0.98)
... very used to CNJ-PST-clean-3.PL and used to they smoked it

(92) ... âhci piko pêyakwan iyikohk **ê-kî-isi-môcikêyihta-mihk**. (CMBK-3-2 271; 0.98)
... still only similar until CNJ-PST-thus-UNSPEC

(93) ê-nôhtêhkatêt ana wîst ôm ê- ê-wâpamiko- ê-wâpahtahk ôma wiyâs **ê-nôhtê-mîci-t**
s/he is hungry this one he too and ê- ê-wâpamiko- s/he shows it that meat CNJ-want-eat-3.SG

... (C4MF 68; 0.97)
...

'That one is hungry, too, and it sees this meat and **wants to eat it** ...'

All of the Conjunct VTI exmplars for this alternation were past tense, except for (93); further, the majority of these top exemplars (that is, examples (89), (90) and (92)) contained the discourse preverb {isi-}.

Transitive Animate Verbs

Independent

(94) ... môy êkw êkonik mîna **ki-kî-wîh-â-wak** ... (AL 1284; 0.91)
... no and those and 2-PST-rely.on-DIR-2SG.3PL

'... now you can't even **rely on them** ...' (Bear et al., 1992, 342-243)

(95) â, **kit-ayâw-â-wak** cî (AL 106; 0.86)
ah 2-have-DIR-2SG.3SG Q

'Ah, do **you have any of that**?' (Bear et al., 1992, 250-251)

(96) ... sapiko mân êkosi **nit-it-â-wak** nôsisimak ... (CMBK-4-2 304; 0.85)
... actually used to so 1-say-DIR-1SG.3PL my grandchildren ...

(97) ... **itâ-wak** mân ôki niwâhkômâkanak ... (EM 160; 0.77)
... say-DIR.3PL.OBV used to FOC my relatives ...

(98) ... ômisi mâna **ni-kî-itâ-wak** nitawâsimisak ... (EM 66; 0.73)
... thus used to 1-PST-say-DIR.1SG.3PL my children ...

'I used to tell my children as follows ...' (Minde, 1997, 36)

Three of the five top exemplars (examples (94), (96), (98)) were forms of *itêw*, 's/he speaks/tells about someone). In each of the five VTA exemplars, the target verbs were the main, and in fact *only*, verbs in their clauses. Estimated probabilities ranged form 0.73 to 0.91.

Conjunct

(99) îh, êwako anima êsa kayâs êkosi **ê-kî-pê-isi-kakêskim-â-cik**
look, this that apparently long ago so CNJ-PST-come-thus-counsel-DIR-3PL.3OBV

otôsk-âyimiwâwa ... (SW 140 0.98)
young people ...

'Look, in this wise long ago did **they use to counsel their young people** ...' (Whitecalf, 1993, 76-77)

(100) ... wâposwa ê-kî-nipahât **ê-wî-kakwê-asam-iko-yâhk** wiya ... (C8GB 13; 0.97)
... rabbits s/he kills him CNJ-FUT.VOL-try-feed-INV-3SG.1PL for ...

(101) ... kita-wâpamikot, **ê-pê-minihkwât-â-yit**, itwêw. (VDC2 485-486; 0.97)
... looking at him CNJ-come-drink-DIR-3.OBV he says

'... looking at him, he said, **to trade it for a drink** ...' (Vandall and Douquette, 1987, 68-69)

(102) ... âta tâpiskôc êkâya kîkway wiyasiwêwin wîyawâw **ê-ohci-tâwiskâ-ko-cik**,
... though like nothing law they CNJ-NEG.PST-be.subject-INV-3PL.3OBV

nânitaw itinikêtwâwi. (C8GB 232-234; 0.96)
something bad when they act thus

'... even though it looked as if **they were not subject to any formal law** when they did do something

wrong ...' (Vandall and Douquette, 1987, 50-51).

(103) miton êsa mân êkotê **ê-kî-isi-sôhkêpit-iko-cik** (cmbk-5-2 72; 0.96)
very apparently used to over there CNJ-PST-thus-support-INV-3OBV.3PL

The majority of the exemplars for the ê-Conjunct Order in this alternation were the same found in the Conjunct outcome in the more Independent vs. Conjunct, with the notable exception of (103). This is likely due to the fact that the majority of Conjunct forms are, in fact, ê-Conjuncts.

5.5.3 Conjunct Type

In the Conjunct type alternation, it does not make sense to analyze both outcomes, as one is simply an *other* case. As such, the exemplars here will only be given for the positive case, (e.g. ê-Conjunct, kâ-Conjunct, or Other-Conjunct).

ê-Conjunct

Inanimate Intransitive Verbs

(104) ... namôy êtikwê **ê-miywâsi-k** ôma ta-nipahtâkêhk. (CMBK-5-2 87; 0.95)
... not apparently CNJ-be.nice-3.SG FOC someone who kills

There was only one correctly identified ê-Conjunct exemplar available, and in this case it was as a main verb of a clause. It's estimated probability is high at 0.95.

Animate Intransitive Verbs

(105) ... môy tâpwê **ê-ohci-ma-miyomahciho-t** ... (CMBK-4-2 159; 0.98)
... not truly CNJ-NEG.PST-RDPLW-feel.well-3.SG ...

(106) ... êkos ânima **ê-isi-tâpwê-t** êwako. (JK 160; 0.97)
... so that CNJ-speak.truth-3.SG this

(107) ... ê-wîcêwâyâhk âskaw ê- **ê-papâmi-mawiso-t** ... (EM 36-37; 0.97)
... we get along with him sometimes ê- CNJ-about-pick.berries-3.SG ...

(108) **ê-papâmi-pa-pêyako-yân** in the spruce -˜ mâka mîn âsay nitâkayâsîmon. (AL 148-149; 0.97)
CNJ-about-RDPLW-be.alone-1.SG in the spruce -˜ but and already I speak English

'**I'd be going about alone** in the spruce ˜ and I'm already speaking English again ...' (Bear et al., 1992, 254-255).

(109) êkos ôma nika-mâc-âcimon nîsta, tânisi **ê-isi-ka-kiskisi-yân** ... (CMBK-1-2 14; 0.97)
so this I will tell bad news I too, how CNJ-thus-RDPLW-remeber-1.SG ...

In the ê-Conjunct outcome for the VAI, all of the top five VAI exemplars make use of preverbs. Interestingly, two of the top five exemplars used the position preverb, {papâmi-} (indicating an action is done throughout an area). Unlike the usual use of a position preverb, {ohci-} in a metaphorical sense, {papâmi-} is used here to actually impart information about spatial position. Another two made use of the discourse preverb {isi-}. Beyond this, the actual semantic criteria of the verbs does not form a cohesive class in this outcome. The estimated probability was high for this outcome, ranging from 0.97-0.98.

Transitive Inanimate Verbs

(110) môy wîhkât nânitaw **ê-ohci-itêyihta-mâhk** ... (CMBK-3-2 162; 1.00)
not ever simply CNJ-NEG.FAST-think-1PL ...

(111) êkwa awa nisîmis, anita wiy êkwa **ê-ohci-nitohta-hk** wîkiwâhk ...
and this my younger sister, there for and CNJ-NEG.PST-listen-3.SG...

(CMBK-4-2 29; 0.98)

(112) ... tâpiskôc namôya kîkway **ê-itêyihta-hkik** onêhiyâwiniwâw. (VDC2 20-22; 0.97)
... Just like nothing CNJ-think-3.PL their Cree way.

(113) â, êkos ê-itihtahk anima, «sâncikilôs [sic]» **ê-itêyihta-hk**, «in the cross»
yes, so he hears thus this, «sâncikilôs [sic]» CNJ-think-3.SG, «in the cross»

ê-itwêwiht. (AA 191; 0.97)
he make such a noise

'Yes, that is what he heard, **interpreting it** as 'sâncikilôs' when they said 'in the cross." (Ahenakew, 2000, 124-125)

(114) ... êkâya kîkway ê-pakitinamâkoyahk, tânisi **ê-itêyiht-am-ahk**. (VDC2 114-115; 0.90)
not what he allows us, what how CNJ-think-THM-21.PL

'... they do not allow us to **think for ourselves** ...' (Vandall and Douquette, 1987, 42-43).

Of the top five exemplars for VTIs, only one (example (111)) was *not* a form of *itêyihtam*, 's/he thinks it'. Interestingly, this exception (a form of *nitohtam*, 's/he listens') is still a sensory verb, which in the VTI falls under the same umbrella as thinking verbs, TI-nonaction. Estimated probabilities were quite high, ranging from 0.90 to 1.00.

Transitive Animate Verbs

(115) ... ma kîkway wîhkâc **ê-ohci-pakitin-i-cik** aniki nikosis Randy ... (CMBK-2-2 43; 0.97)
... not what ever CNJ-NEG.PST-let.go-INV-3PL.1SG those my son Randy ...

(116) ... môy âhpô ê-ohci-kiskêyim-ak awa kâ-wî-wîkimak awa Tommy, môy
... not even I did not know him this whom I am going to live with this Tommy, not

ê-ohci-kiskêyim-ak. (CMBK-4-2 114; 0.97)
CNJ-NEG.PST-live.with-DIR.1SG.3SG

(117) ... tânsi ê-isi-sîhkimicik, tânisi **ê-isi-nitawêyim-i-cik**, nikî-tôtên. (EM 92; 0.96)
... how he urges me, how CNJ-thus-want-INV-3PL.1SG I do it

'... what they urged me, **what they wanted me to do**, I would do.' (Minde, 1997, 66-67)

(118) êwakw ânima kêhcinâ aya ê-kî-miywêyihtamân, ê-kî-oh- aya **ê-kî-isi-wâpam-ak**
this that certainly well I was glad ê-kî-oh- well CNJ-PST-thus-see-DIR.1SG.3SG

niwîkimâkan ôtê kâ-pê-wîcêwak ... (EM 65; 0.93)
my husband over here I come to marry him

'I certainly used to be happy that **I could see my husband** in this light when I came over here to be married to him ...' (Minde, 1997, 36-37)

(119) ... wiy âh-apisîs piko **ê-kî-asam-ikawi-yâhk**. (CMBK-1-2 25; 0.92)
... for very small a bit CNJ-PST-feed-UNSPEC-1.PL

Although the VTA exemplars for the ê-Conjunct outcome have little cohesion, all but (116) represent a past action, even if not represented in the morphology. Beyond this,

ê-Conjunct exemplars often contain first person goals, as in (115), (117), (119). These exemplars ranged from 0.92 to 0.97.

kâ-Conjunct

Inanimate Intransitive Verb

(120) ... ita êsa mân êtikwê ê-kî-osâpit, **kâ-kîsikâ-yik** ... (CMBK-5-2 57; 0.75)
... there apparently used to apparently he watched from there, CNJ-DAY-3.OBV

(121) ... wiya pîhc-âyihk kâ- **kâ-pipoh-k** kâ-kî-ayâyâhk. (C2GB 18; 0.63)
... for inside kâ- CNJ-snow-3SG we were

The kâ-Conjunct outcome for the VII had only two valid exemplars. Each of these exemplars were used not prototypical verbs, instead referring to refer to temporal periods. In (120) the exemplar, a form of *kîsikâw* is used as a temporal prepositional phrase/adjunct. In (121), the verb *kâ-pipohk* ('it is winter'), is used nominally to simply mean 'winter'. These exemplars were not as well predicted as those covered previously, ranging from 0.63 to 0.75.

Animate Intransitive Verb

(122) ... mîn êkâ awiyak **kâ-kî-minaho-t**, âhci piko pêyakwan ê-miyiht wiyâs ...
... and no someone CNJ-PST-hunt-3.SG, still a bit one he gives to me meat ...

(CMBK-4-2 264; 0.73)

(123) ... êkwa **kâ-minaho-cik** ôkik nâpêwak ... (C2GB 14; 0.72)
... and CNJ-hunt-3.PL these men ...

(124) cikêmô kikî-miyikonaw kôhtâwînaw, kîstanaw **kâ-nêhiyâwi-yahk** ... (JK 7; 0.63)
of course we were given it by him our father we too CNJ-be.Cree-21.PL ...

(125) ... misatimwak ê-têhtapiyâhk, itê **kâ-minaho-cik** nôhtâwînânak. (CMBK-4-2 250; 0.59)
.. horses we ride there **they hunt** CNJ-hunt-3.PL

(126) êkwa mîna pikw îta **kâ-pîhtikwê-yan** ê-mîcisoyan ... (AL 71; 0.57)
and also only there CNJ-come.in-2.SG you eat

Three of the five exemplars ((122), (123), and (125)) for the VAIs in the kâ-Conjunct outcome concern forms of the lexeme *minahow* ('s/he hunts/kills'). Beyond this, this is little that can be generalized about these exemplars. Estimated probabilities were relatively low, ranging from 0.57 on the low end to only 0.73 on the high end.

Transitive Inanimate Verb

(127) ... wâhyaw ôm ôma **kâ-it-am-ân**, môy âhpô nikiskisin tânis ânim ê-isiyîhkâtahkik
... far away then the fact that CNJ-call-THM-1.SG, not even I know what that they call it

... (CMBK-1-2 237; 0.91)
...

(128) ... 'iyisâhowin' anima ka- **kâ-it-am-ihk** aya ... (EM 75; 0.83)
... 'iyisâhowin' this ka- CNJ-call-THM-UNSPEC this one ...

'... "resisting tempatation" as **they would call it** ...' (Minde, 1997, 46-47)

(129) êkoyikohk isko ê-kî-nôhtê-âcimostawak awa niwîcêwâkan, êwak ôm ôma
only then until they wanted to tell a story FOC my spouse, this one FOC that

kâ-nitawêyiht-ahk. (CMBK-3-2 48; 0.66)
CNJ-want-3.SG

(130) ... anima **kâ-nôhtê-kiskêyiht-ahk** nâha, êwako ê-kî-pawâmit anima ... (SW 39; 0.61)
... that CNJ-want-know-3.SG that one, this one he had a dream spirit that ...

'... what that one **wants to know about**, that the woman had a dream spirit ...' (Whitecalf, 1993, 36-37)

(131) êwakw ânim ânohc **kâ-mâmiskôt-ahk** ayamihêwiyiniw ... (EM 78; 0.61)
this that today CNJ-talk.about-3.SG priest ...

'his is what the priest **talked about** today ...' (Minde, 1997, 52-53)

The VTI exemplars were made up mostly of verbs of speech. In (127) and (128) the exemplar verb *itam* ('s/he calls it so') is used; in (131) the verb *mâmiskôtam* ('s/he talks about it') is used; and finally in (129) where the exemplar *ê-kî-nôhtê-âcimostawak* refers to telling a story. The remaining exemplar verb, *kâ-nôhtê-kiskêyihtahk*, refers to knowing. This is strikingly similar to As in the previous outcome, all these exemplars fall

under the banner of TI-nonaction. This class of verbs had a large range in its estimated probabilities, with the lowest exemplar estimated at 0.61 and the highest at 0.91.

Transitive Animate Verb

(132) ... mâk êkwa awa **kâ-pê-wîhtamaw-it** nitôsim ... (CMBK-4-2 19; 0.71)
... but and this CNJ-come-tell.about-INV.3SG.1SG my stepson ...

(133) ... kîkway ôki **kâ-wîhtamaw-icik** nitawâsimisak ... (CMBK-4-2 202; 0.71)
... what these CNJ-tell.about-INV.3PL.1SG my children ...

(134) êwakw âwa **kâ-wîhtamaw-ak** anohc ... (JK 4; 0.64)
that one that CNJ-tell.about-1SG.3SG today ...

The VTA kâ-Conjunct exemplars were fewer in number than the VAI and VTI classes with only three valid exemplars present. Similar to what was seen in the VTI class, each exemplar is a form of a speech verb. Probability estimates ranged from 0.64 to 0.71.

Other-Conjunct

Intransitive Inanimate Verbs

(135) ... môniyâw ê-pêhtât nêtê **ta-takopay-iyiki** anihi. (CMBK-3-2 134; 0.53)
... white man he waits there CNJ-arrive.3.PL.OBV FOC

In the final outcome, the Other-Conjunct, the VII had only a single valid exemplar. Here the exemplar is a subjunctive form acting as a temporal adjunct. The probability estimate for the one exemplar was quite low at 0.53.

Animate Intransitive Verbs

(136) ... tânisi k-êtôtamân, **mêstohtê-yêko** pê-miyikawiyâni wêpinâson ... (JK 160; 0.89)
... what I will do, die-2PL.CNJ.FUT.COND if I am given it cloth

'... what will I do **when you are all gone** if someone comes and gives me cloth ...' (Kâ-Nîpitêhtêw, 1998, 132-133)

(137) ... ahpô kikaskihtân **ta-nipâ-yan** ... (SW 112; 0.67)
... or you are able to CNJ-die-2.SG ...

(138) ... êkwa awiyak nôhtê-papâmitâpâsoci, ta- **ta-papâmitâpâso-hk** ... (AL 42; 0.63)
... and someone when he wants to ride about. ta- CNJ-ride.around-UNSPEC

(139) misawâc ôta, ispî mêht- [sic] **mêstohtê-twâwi** ... (JK 18; 0.53)
in any way here, when mêht- [sic] die.off-3PL.FUT.COND

'In any case, **when all those here will have died** ...' (Kâ-Nîpitêhtêw, 1998, 64-65)

All four valid VAI exemplars were hypothetical, time dependent, verbs. In most cases, the exemplars were in the subjunctive Conjunct/future conditional form, though even when simply in the ka/ta-Conjunct (as in (138)) the conditional meaning is still present. Expected probabilities had quite a large range, from 0.53 to 0.89.

Transitive Inanimate Verbs

(140) ... piko kâwi **ka-kîwêtot-am-ahk** k-âtoskêyahk, ka-kakwê-pimâcihoyahk ... (EM 96; 0.85)
... a bit again CNJ-return-THM-21.PL we work, we try to make a living

'... so **we will have to go back** and work to try and make a living ...' (Minde, 1997, 72-73)

(141) mistahi ka-miywâsin, êwak ôma kîstawâw, **ka-kiskinowâpaht-am-êk** ôma
very it is good, this the you all too, CNJ-learn.by.watching-THM-3.PL this

kâ-wî-isîhcikêyâhk oskinîkiskwêwak, kwayask ... (JK 158; 0.84)
we are going to do it young women, properly

'It will be very good for you too, the young women, t**o watch what we are going to do and learn from it** ...' (Kâ-Nîpitêhtêw, 1998, 130-131)

(142) ... wiya kiyânaw kikaskihtânaw kîkway ka-kî-nipahtamâsoyahk, kayâsi-pimâcihowin
... for us we can what we killed, old way of life

ka-otin-am-ahk ... (C2GB 45; 0.74)
CNJ-take-THM-21.PL ...

'... for we are able to kill things for ourselves and **to take up** our traditional way of life ...' (Bear et al., 1992, 82-83)

(143) môy pikw êkosi k-êsi-mâmitonêyihtamahk, **ka-tôt-am-ahk** anima ... (EM 76; 0.78)
no only so we should think that way, CNJ-do-THM-21.PL that ...

'We should not only think that way, **we should do it** ...' (Minde, 1997, 48-49)

(144) ... ê-miyohtwât an[a] îskwêw ê-wîcihât anih ôskinîkiwa, **ta-pônihtâ-yit**
... she is good natured that woman she helps him that young man, CNJ-quit-3.OBV

minihkwêwin ... (EM 134; 0.76)
alcohol ...

'... that woman is good-natured and helps that young man **to quit drinking** ...' (Mindc, 1997, 92-93)

The VTI class has five exemplars, all of which occur in the ka/ta-Conjunct. In most cases, these are translated as infinitive forms and nearly always act as non-main verbs. For the VTIs, probabilities estimates were moderate, ranging from 0.76 and 0.85

Transitive Animate Verbs

(145) ... âta kâ-nisitohtahkik, âta **kitota-twâwi**, tâpiskôc êkâya
... although they understand, although speak-2PL.3PL.FUT.COND, for instance not

ê-pêhtâskik (VDC2 19-20; 0.84)
they hear you ...

(146) k-âyimômâyahk kîc-âyisiyinînaw, ahpô **ka-pâhpih-â-yahk** ê-kitimâkinâkosit ...
when we gossip about him our fellow man, or laugh-THM.21PL.3PL he is pitiable ...

(JK 9; 0.84)

'When we gossip about our fellow man, or **if we were to laugh at someone** who looks pitiable ...'

(Kâ-Nîpitêhtêw, 1998, 54-55)

The final class, the VTAs, has only two valid exemplars. Both are conditional verbs, though only (145) had the verb in the subjunctive Conjunct form. Instead, (146) contains *ka-pâhpihâyahk*, 'if we laugh at him', in the ka/ta-Conjunct without any particle that might suggest conditionality. It is also worth noting that in (145) the exemplar verb is, as has been seen previously in many instances, one of speech. Both exemplars were well predicted, with probability estimates of 0.84.

Chapter 6

Discussion

This chapter will discuss the results presented in the previous chapter and the ways in which they inform our understanding of Order and alternation in Nêhiyawêwin. The first section will discuss the behavioural profiles (that is, the set of "elements co-occurring with a word within the confines of a simple clause or sentence in actual speech and writing" as defined by Gries and Divjak (2009, 63)) suggested by the results and how these relate to previous research on Order. The next section will then discuss the statistical veracity of the logistic models. Following this, is a brief discussion of how the extraction of exemplars is the previous chapter can inform our understanding of Order. Finally, the chapter will close with a summary of what this research has taught us about Nêhiyawêwin Order.

6.1 Independent vs. Conjunct

In the alternation between the Independent and the general Conjunct, the majority of significant effects, regardless of verb class, were predictive of a Conjunct form, rather than the Independent. A general summation of the effects across verb classes is given in Table 6.1. The fixed effects are split into four main categories: effects of actors, effects of goals, effects of preverbs, and effects representing semantic classes. For a discussion

Table 6.1: Multivariate Effects: Independent vs. Conjunct. Each cell is labeled with the outcomes for which an effect is significant. If a cell is coloured green, the effect increased the likelihood of an Independent form, while a red cell represents an effect increasing the likelihood of a Conjunct form

	Effects	II	AI	TI	TA
Actor	actor.obv		CNJ		CNJ
	actor.sg		CNJ		CNJ
	NA.persons.actor			IND	
	actor.1				IND
	actor.2			IND	
	actor.3			CNJ	
Goal	NI.Place.goal			CNJ	
	NDIbody.goal			CNJ	
	NI.nominal.goal			CNJ	
	NA.persons.goal				CNJ
	goal.obv				CNJ
	px1sg.goal			IND	
	goal.2				CNJ
Preverb	PV.Time	CNJ	IND		CNJ
	PV.Discourse		CNJ	CNJ	
	PV.Position				CNJ
	PV.Move				CNJ
Semantic class	Food				CNJ
	Do			CNJ	
	Money/count			CNJ	

of how the random effect of lexical identity affects Order choice, see section 5.4.4 in Chapter 5.

Preverbs seemed to only increase the likelihood of a Conjunct form, with one exception. This behaviour may suggest that the Conjunct is a more modified category. In particular, preverbs of discourse suggest a verb that is not simply declarative in structure, providing some information about the discourse, either by expressing uncertainty and hesitation (as in {ayi-}, {ata-}, both hesitation/planning markers like the English *um*), or connection to the manner in which verb is performed (as in {isi-}, meaning 'it is done thusly,' or {isko-}, 'it is done to such an extent'). This sort of behaviour conforms with

the descriptions of Cook (2008, 162), who suggests that the Conjunct to be more likely in a medial context; that is, it is more likely to not be at the start of a conversation or story. This description implies the Conjunct order is somehow related to the discourse structure of the utterance. The only preverbs associated with the Independent is in the VAIs, where preverbs of time increase the likelihood of the order. This is peculiar for two reasons: Firstly, preverbs of time include PV.ka, an irrealis preverb (usually interpreted as a future definite form in the Independent) that is also present in all ka-Conjunct forms, which make up a large amount of the Other-Conjunct class. Following from the first peculiarity, the second is in the disagreement between the VAIs and the VTIs and VTAs in the direction of the effect of PV.Time. Although no single effect was significant in all classes, no effect other than PV.Time differed in its direction of association throughout the inflectional classes.

Actor persons were not significant for all classes, but when present, local actors increased the likelihood of an Independent, while third person actors increased the likelihood of a Conjunct. Also interesting is the distribution of Independent effects across conjugation classes: the VTI had three Independent effects, while the VAI and VTA had only one each. The VII had only one significant effect for Independent forms, that of PV.Time, though this particular lack of effects is likely due to the lack of tokens in analysis. Also relating to the Independent is the fact that Independent effects were almost always argument effects, such as actor.1.

A final note, and one that affects all classes, is that PV.Position includes the preverb *ohci-*, which can mean 'from', but is also used as a negative past marker, as previously discussed. According to Proulx (1991, 146), {ohci-} in Swampy Cree (which is analogous to the Nêhiyawewin preverb of the same form) derives its negative meaning from the original spacial interpretation going back to Proto-Algic; According to Proulx (1991), this exemplifies Filmore's *Moving world metaphor* wherein time is a continuous process and the all things move forward through it (Proulx, 1991, 146). In fact, the majority of

instances of the {ohci-} preverb seem to act in this manner rather than in the more literal spacial sense, and roughly 75% of all PV.Position tokens were {ohci-}.

Overall, the alternation between the Independent suggests that the Conjunct is a more marked class, and one that is more associated with modifying preverbs, especially of those of time and discourse.

6.2 Independent vs. ê-Conjunct

The pattern of effects is significantly different for the Independent vs. ê-Conjunct alternation than for the Independent vs. Conjunct alternation. This difference suggests a difference in the *type* of alternation. The effects of this alternation are detailed in Table 6.2.

Similar to the previous alternation, local actors, when significant, always increased the likelihood of an Independent Order, while third person actors increased the likelihood of the ê-Conjunct order for VTIs; obviative actors did the same, but in the VTA class. The presence of an overt actor representing a person also increased the likelihood of the Independent, though only significantly for the VTIs. Together, these effects suggest an Independent order that is associated with a higher position in the Nêhiyawêwin person hierarchy (reproduced in (147)), while the ê-Conjunct generally associated positions on the lower level of the hierarchy (i.e. non-local participants).

(147) $2 > 1 >$ Unspecified Actor $> 3 > 3' > 3''$

As previously, the Independent was generally associated with effects dealing with actors and goals, with the exception of preverbs of time. Unlike the previous alternation, preverbs of time significantly increased the likelihood of the Independent for both the VIIs and the VAIs, though it is still associated with the ê-Conjunct in the VTA class. Other than this set, all other preverb effects which were significant were associated with the ê-Conjunct outcome. It is worth noting, that preverb effects were mostly significant

Table 6.2: Multivariate Effects: Independent vs. ê-Conjunct. Each cell is labeled with the outcomes for which an effect is significant. If a cell is coloured green, the effect increased the likelihood of an Independent form, while a red cell represents an effect increasing the likelihood of an ê-Conjunct form

	Effects	II	AI	TI	TA
Actor	actor.1		IND		IND
	actor.2		IND	IND	IND
	actor.3			ê-CNJ	
	actor.obv				ê-CNJ
	NA.persons.actor			IND	
Goal	NI.place.goal			ê-CNJ	
	Px1Sg.goal			IND	
Preverb	PV.Discourse		ê-CNJ	ê-CNJ	ê-CNJ
	PV.Move				ê-CNJ
	PV.Position				ê-CNJ
	PV.Time	IND	IND		ê-CNJ
Semantic class	Do			ê-CNJ	
	Food				ê-CNJ
	Money.Count			ê-CNJ	

only for the VTA class. In fact, only preverbs of discourse and preverbs of time had effects in any other verb class. This suggests, as in the Independent vs. Conjunct alternation, that the ê-Conjunct is a form that is more marked/altered (except for time). Finally, semantic classes were again only significant when influencing the ê-Conjunct, and even then only in transitive classes.

6.3 Conjunct Type

The Conjunct Type alternation was significantly less 'cohesive'. That is, less can be said about an outcome across verb classes. As can be seen in Table 6.3, even when an effect is present in multiple verb classes, it is not always the case that the effect increased or decreased the likelihood of the same outcome in each conjugation class. (e.g. while PV.Discourse is significant for VAIs and VTAs, the effect increases the likelihood of

ê-Conjunct in both, but decreases the likelihood of kâ-Conjunct in the VAIs and Other-Conjunct in the VTAs. Similarly PV.Position, increased the likelihood of the ê-Conjunct in the VAIs, VTIs and VTAs, but it decreases the likelihood of kâ-Conjunct in VAIs and VTAs while decreasing the likelihood of the Other-Conjunct in the VTIs.

Table 6.3: Multivariate Effects: Conjunct Type. Each cell is labeled with the outcomes for which an effect is signficant. If a cell is coloured green, the effect increased that outcome, while a red cell represents an effect decreasing likelihood.

		II	AI	TI	TA
Actor	actor.1		ê- kâ-		ê- Other-
	actor.2		ê- Other-	ê- Other-	ê- Other-
	actor.3		ê- kâ-		
	NA.persons.actor			kâ-	
	NDA.Relations.actor		kâ- Other-		
	Pl.actor		kâ-		
	Prox.actor			ê-	ê- kâ-
	Sg.actor		ê- kâ-		
	NI.nominal.goal			ê- Other-	
	NI.object.goal			kâ-	
	Med.goal			Other-	
	Sg.goal			ê-	ê- kâ-
	Prox.goal			kâ-	
	Px1Sg.goal				ê-
Preverb	PV.Discourse		ê- kâ-		ê- Other-
	PV.Move				
	PV.Position		ê- kâ-	ê- kâ- Other-	ê- kâ-
	PV.Qual		Other-		
	PV.WantCan		kâ-	ê- kâ-	
Semantic class	Cognitive				kâ- Other-
	Cooking		kâ-		
	Health		Other-		
	Speech			ê- kâ-	
	Weather	kâ-			
Reduplication	RdplW		ê-		

In all classes excluding the VII, preverbs of position increased the likelihood of the ê-Conjunct. Conversely, the ê-Conjunct's likelihood was decreased by the presence of second person actors while the Other-Conjunct was increased for the same variable. First person actors significantly increased the likelihood of the ê-Conjunct, but only for the VAIs and VTAs. Third person actors also increased the likelihood ê-Conjunct, but only significantly for the VAIs.

Perhaps most clear is the effect of II.weather on the kâ-Conjunct. The use of a Conjunct form for weather verbs seems to allow for the use of the verb as a durative state as in (148), where the verb *kâ-pipohk* is used to mean 'during winter', rather than being used as a more declarative statement.

(148) ... awâsisak wâwîs **kâ-pipohk**
... children especially **when it is winter**

'... especially for children in winter.' (Minde, 1997, 137)

6.4 Model Statistics

Modelling of the alternation varied in efficacy. With the main mixed effects models, in the Independent vs. Conjunct alternation only the VAI and VTA models were quantitatively well fit. This may simply be due to frequency effects. The Independent vs. ê-Conjucnt alternation showed more success in model fitting: only the VIIs failed to produces a relatively well fit model. Again, the VAI and VTA classes had the highest ρ^2 scores. The Conjunct type alternation faired even better, with each class exhibiting well fit models. Assuming this is not due to data-specific effects, this may suggest that the alternations between the Conjunct Types and the Independent vs. ê-Conjunct are more linked to the sorts of effects used in this dissertation (primarily morphological and semantic effects).

In contrast with a number of previous studies, which conclude that, while morphosyntactic effects provide some amount of predictive power, it is the semantic effects that are most predictive, this dissertation finds that semantic and morphological effects are similarly effective as predictors. For example, in Arppe (2008) semantic classifications of the arguments of the predicates, as well as semantic classifications of the predicate verb phrases themselves, were found to be better predictors in the alternation between various verbs of thought; Divjak and Arppe (2013) similarly highlights the importance of semantic classification of subjects and infinitives when

choosing between verbs of trying in Russian; and Abdulrahim (2013), who found that semantic classifications can be useful in profiling *go* and *come* verbs in Arabic (cf. the results of Klavan (2012), reports that semantic effects are less effective than morphosyntactic variables for accounting for the alternation between the use of adessive case and the use of the adposition *peal* (meaning 'on') in Estonian). Although the findings of this dissertation varied by alternation type and conjugation class, overall the difference between semantic and morphological effects was slight for the Nêhiyawêwin Order alternation.

The reason for the discrepancy between the results of this dissertation and those mentioned above could be due to differences in the ways semantic classes were defined, the differences between the languages being studied, or even the specific types of linguistic phenomena that are studied. While Finnish and Russian are synthetic languages with rich case systems, their verbal systems are not as morphologically complex, lacking a wide system of preverbs or polypersonal agreement, for example. In fact, none of the above mentioned languages' verb systems match the morphological complexity of Nêhiyawêwin's. It is possible that Nêhiyawêwin's polysynthesis bolsters the explanatory power for morphological effects in modelling these alternations. Additionally, the nature of the alternations being studied in the above studies are significantly different than those in this dissertation. While Arppe (2008), Divjak and Arppe (2013), and Abdulrahim (2013) focused on an alternation between near synonymous terms generally to do with a single semantic domain (*thinking*, *trying*, and *motion* verbs respectively), the alternation of Order is more structural, as described in Chapter 2. Only Klavan (2012) could be said to be somewhat structural, comparing the use of an adposition with a case marker. Finally, the research in this dissertation differs from previous research like Arppe (2008), Divjak and Arppe (2013), and Abdulrahim (2013) methodologically in that the models used in this research make use of mixed effects models. Although it is not immediately obvious how controlling for the random effect of lemma identity would affect the usefulness of

semantic and morphological predictors, its inclusion is significant. Interestingly, Klavan (2012) *did* make use of mixed-effect models, and their results were most similar to those of this dissertation regarding the importance of semantic effects. Despite the differences between this dissertation's research and those previously mentioned, all studies demonstrate the ability to model alternations in very divergent languages using logistic regression modelling.

6.5 Exemplar Extraction

Exemplar extraction was mostly successful, and useful exemplars were presented in Chapter 5. Interestingly, in actually clustering tokens in order to extract exemplars, a pattern emerged. As a reminder, the purpose of clustering tokens was to avoid over-representation of a few items which all contained the same combination of strongly predictive variables . Despite the clustering work done, in some cases this still occurred. For example, nearly all properly predicted (and with high estimated probabilities) Independent VAIs, regardless of which alternation they were seen in, were a form of the lemma *itwêw*. Although one could simply select only one such exemplar for any given lemma, there are a number of issues in doing so. Most importantly, doing so ignores the fact that the estimated probabilities for *itwêw* are significantly higher than other lemmas (e.g., in the Independent vs. Conjunct alternation, the lowest estimation for an *itwêw* lemma is 73%, while the next highest rated token of a different lemme is a full 10 percent-points lower). By choosing to ignore all but one *itwêw* lemma, one would be highlighting less certain predictions. Further, it is also worth highlighting the fact that VAI quotatives are closely related to the Independent.

6.6 Summary

Motivating features tended to vary between alternations and inflectional classes. What can be said, however, is that when alternating with the Independent, the Conjunct order as a whole tended to be a more marked form, and one that was often associated with preverbs that gave information about 'position' and 'discourse.' Interestingly, some of the most numerous preverbs from these classes {isi-}, {ohci-}, and {isko-} are derived from what Wolfart calls 'relative roots,' which require an antecedent (Wolfart, 1973). According to Cook (2014, 55), this closed class has only one member beyond the previous three morphs mentioned: {tahto-} 'as many as.'[1] This behaviour comports with the description from Cook (2014) who claims the Conjunct to be one that acts anaphorically.

Specifically investigating the alternation between the Independent and the ê-Conjunct, this dissertation finds a similar result to the general Independent vs. Conjunct alternation, suggesting that the ê-Conjunct is a form used when referring to actions involving non-speech act participants, and especially when using a word that has a referent elsewhere in the discourse.

Despite the relatively well fit models of the Conjunct Type alternation, there was little in terms of a coherent pattern. In nearly all cases, no predictor was significant for more than half of the inflectional classes. The two exceptions to this are the predictors of PV.Position, which was significant for the VAIs, VTIs and VTAs and always increased the likelihood of an ê-Conjunct; and actor.2, which was significant for the VAIs, VTIs and VTAs and always increased the likelihood of an Other-Conjunct form and always decreased the likelihood of the ê-Conjunct. This means that when speaking directly to a person, the Other-Conjunct forms are more likely to be used, perhaps due to hypothetical statements that one might make when counselling someone or giving advice, as in *pimohtêyani* 'if you walk...'. Further, the overall less clear picture is, perhaps,

[1]This preverb was included in the PV.Quantity class due to its straightforwardly quantifier-like semantics.

understandable as the Conjunct types are much more clearly associated with certain semantic interpretations (e.g. the subjunctive Conjunct always produces a hypothetical statement).

Finally, it appears that, although mixed effects models were always the best fit models over all, individual inflectional classes sometimes showed other models out-performing, as the Lemma-as-random-effect model (LEM) does in Table 5.49, reproduced as 6.4 here.

Table 6.4: Model Comparisons. Independent vs. Conjunct, bold items represent a very good model fit, per McFadden (1973)

	VII		VAI		VTI		VTA	
	τ	ρ^2	τ	ρ^2	τ	ρ^2	τ	ρ^2
Sem.Morph	0.31	0.05	0.47	0.15	0.32	0.09	0.35	0.10
Lem	0.36	0.12	0.52	**0.26**	0.34	0.15	0.42	0.18
Sem.Lem	0.35	0.12	0.52	**0.27**	0.36	0.15	0.44	0.19
Morph.Lem			0.52	**0.27**	0.39	0.16	0.43	0.19
Mixed Effects	0.35	0.12	0.53	**0.27**	0.42	0.16	0.45	**0.21**

Interestingly, even though Lemma-as-random-effect models were rarely well fit enough or possessing substantial variation explanation, they still often performed well. In 6.4, τ scores were all over 0.30 in the lemma-only model, and ρ^2 values ranged from 0.12 to 0.18. A similar pattern was seen in other alternations. Further, even when other information was included, these scores did not rise markedly. These results suggest that lemma-specific effects offer a substantial, if not holistic, explanation for Order alternation; that is to say, much of the alternation appears to be lexically motivated.

Chapter 7

Conclusion

This dissertation has investigated how one can make use of mixed effects logistic regression to model Order in Nêhiyawêwin as a form of alternation. This technique produces estimates of log odds between two alternate outcomes; that is, it predicts the proportion of how likely two outcomes are to occur based on some context on the long run.

Given the results of Cook (2014), the implicit research question underlying this investigation is if morphosyntactic and semantic features can allow for more accessible description of the difference between the different Order types and subtypes. As the results in Chapter 5 show, this assumption was largely not supported for this specific set of alternations: indeed, the main results of the multivariate modelling, that preverbs of discourse (which regularly connect a verb to some other referent in the discourse) and position (which nearly always actually represented a past tense form), largely confirm the main finding that the Conjunct is associated with anaphora.

Beyond this particular hypothesis, five predictions were proposed. in the following, I will evaluate each of these predictions based on the results described previously.

1. *Overall, modelling will be successful though constrained* (*likely due to the small size of the corpus*).

This prediction proved more or less correct. Although modelling was performed, there were few significant multivariate effects. Further, although modelling provided some insight into how the phenomenon alternates, no substantial and humanly understandable picture of the phenomenon emerged from the modelling.

2. *Due to a lack of syntactic data, mixed effects modelling based on the Nêhiyawêwin corpus will be able to provide some insights, but model fits will rarely achieve a level beyond well fitting.*

   This prediction was ultimately unsupported. Although the Mixed Effects VII and VTI models of the Independent vs. Conjunct alternation, and the VII model in the Independent vs. ê-Conjunct alternation, did not have particularly well fit models, all other Mixed Effects models were relatively well fit. This indicates that the semantic, morphological, and whatever syntactic information was available, was sufficient enough to predict the alternation. Although higher level relationships such as anaphoric reference seemed to be strong predictors, the use of a lemma alone explained substantial, if not quite significant, amounts of variation. This suggest that, while morphological, syntactic, and semantic effects are necessary to define the alternations in this dissertation, lemma identity has some power to predict a particular Order or Order subtype.

3. *Semantic variables will do more to explain variance than morphological variables* (*as in Arppe* (*2008*)).

   This prediction was, in fact, unsupported. Excluding the semantic groupings of preverbs, the majority of the semantic classification information was not significant enough to warrant inclusion as a fixed effect in multivariate modelling. Models without semantic information generally performed as well as those with semantic

information. Further, there was no straightforward patterning of how a semantic class might affect Order choice.

4. *The Conjunct Type alternation will be significantly less cohesive in its results* (*due to the straightforward syntactic/semantic choices driving it, which are not reflected in the variables of the data set*).

   The results of this dissertation conform to this prediction. While models for this alternation were well fit, at least in the mixed effects case, there was not an easily identifiable or human interpretable motivation for the results in the Conjunct Type alternation.

5. *The alternation between the Independent and the ê-Conjunct will be the most robust and well-fit model* (*because the two forms are nearly synonymous in many cases*).

   Most surprisingly, this prediction was unsupported. The Conjunct Type alternation overall had the best fit models, along with all inflectional classes being relatively well fit. On the other hand, in the Independent vs. ê-Conjunct alternation only the VAI, VTI, and VTA classes had scores above the threshold for a good fit. It is not immediately clear why the Conjunct Type alternation was better fit, but it seems possible that this is simply due to the fact that the model almost always guessed the ê-Conjunct regardless context. Given that the ê-Conjunct was simply so much more numerous than other outcomes, this proved a valid strategy.

There were also four identified research questions that were addressed in this research:

1. *Can mixed effects modelling be used in investigating complex morphological phenomenon using a small but richly tagged corpus?*

This research question was answered partially in the affirmative. Although the models were unable to provide clear-cut explanations of Order, the alternations being investigated were able to be partially explained by modelling the morphological, semantic, and syntactic information available.

2. *Are the alternations between the Independent and Conjunct, the Independent and the ê-Conjunct, and the Conjunct types similarly able to be modelled, or are some of these alternations easier to model than others?*

   Different alternation types produced substantially different results. As previously mention, although the Conjunct Type alternation had the most impressive modelling (in terms of classification improvement as compared to baseline and reduction in badness of fit), its results were also the least humanly interpretable with no clear story as to how or why certain predictors interact to predict a particular Conjunct type.

3. *What are the variables that increase the likelihood that a lemma will occur in a particular Order/outcome?*

   Variables varied widely between inflectional classes and alternations. Notably, however, wherever discourse or position preverbs (which, together more accurately represent antecedent reference preverbs), they predicted a Conjunct form over an Independent form. Further, when predicting an Independent vs. either the Conjunct order in general or the ê-Conjunct in specific, any time a second person actor was a significant predictor it predicted an Independent form.

   Together, the results of this dissertation suggest the Independent is usually a more isolated utterance and one that is more associated with speech directed to an interlocutor, while the conjunct is generally one more connected to previous utterance (as suggested by Cook (2014)). However, as seen by the estimated probabilities of exemplars, the choice of Order is not one that is categorical. Instead,

there is some amount of variation inherent to the alternation.

In addressing these hypotheses, predictions, and research questions, this dissertation showed the ways in which a small but richly tagged corpus can be used in traditional quantitative linguistic analysis. Although the results in this dissertation did not produce a straightforward answer as to what motivations exist in choosing between the Independent and Conjunct Order, the Independent and ê-Conjunct, and finally the three different Conjunct types identified in this dissertation, it did demonstrate that one can conceive of Order as a form of alternation. Further, this research, based on the work of Arppe (2008), Klavan (2012), Abdulrahim (2013), and Divjak and Arppe (2013), among others, showed how one can make use of parametric multivariate and mixed effect modelling to help understand the phenomena. Specifically, this research confirmed that, when deciding between and Independent and Conjunct form, the Conjunct form is more likely if a verb contains a preverb representing a high-level/anaphoric relationship. It also demonstrated that Nêhiyawêwin morphosyntactic features are less covariate with higher level discourse structures and that Order is not as highly based in semantics as something like choice of synonymous verbs.

Given the results of anaphoric preverbs such as {ohci-}, {isko-}, and {isi-} being associated with the Conjunct, future research in this area should focus primarily on syntactic and pragmatic annotation. Doing so would likely produce much better modelling, especially if combined with the morphosemantic information used in this dissertation. The addition of further corpora, such as the Bloomfield corpus (as used in Schmirler (2022)) would also increase the sort of analyses that could be performed. In terms of the techniques used in analysis, combining logistic regression with neural technologies could improve results. This sort of synergy is seen in Tzougas and Kutzkov (2023), which documents a method to enhance a binary regression model by using a neural network to learn and binarize features from data and then fit a model based on those features with a single hidden-layer neural network. Further, if much of this

alternation rests in higher level phenomena and contextual use in the discourse, and if context is something negotiated between the interlocutors (as conceived of by Hirst (1997)), richer tagging that provides some level of discourse analysis would likely allow for more comprehensive analyses. Ideally, the inclusion of more sophisticated techniques and a higher level of annotation will result in a well-rounded explanation of Order. It is also worth considering the use of psycholinguistic experimentation with native speakers to investigate Order.[1] As Klavan (2012) and Arppe and Järvikivi (2007) demonstrate, the combination of multiple types of evidence and methodologies can improve the understanding of a given phenomenon. Of course, experimentation with Indigenous communities requires careful ethical consideration. Also concerning is the paucity of available native speakers who are willing to participate in such activities. In fact, a major motivation for this dissertation was to inspire hypotheses that researchers can use to best target their research. This way, participants can take part in only those tasks that are most likely to produce informative results. Given the often aging population of Nêhiyawêwin speakers, making the best use of participant time is especially important.

In its original conception, this dissertation was a way to investigate Order as a phenomenon in a systematic and applicable manner. It was my hope that the results from this research would be more easily translatable to pedagogical materials than previous attempts at explaining Order. Over the course of this dissertation, I have confirmed that Order in Nêhiyawêwin is a multifaceted, complex phenomenon that involves multiple levels of linguistic analysis. Moreover, I have demonstrated that, even with powerful technologies, there may not be a clear or simple explanation for how Order operates. This dissertation has primarily dealt with the ways in which quantitative and computational techniques can be used to assess Order in Nêhiyawêwin. While, the quantitative and linguistic analyses in this dissertation may be esoteric to students of Nêhiyawêwin

[1]In fact, I attempted a sentence-completion experiment with native speakers during a dictionary recording session in Maskwacîs, Alberta, generally unrelated to this dissertation. Although all speakers were cooperative, they reported frustration with the task. As a result, further experimentation was put on hold until further insights could be gathered.

without prior knowledge of these topics, it is my hope that the exemplars provided in Chapter 5 may provide some guidance for them. At the very least, by investigating Order from an empirical point of view, the variability of the phenomenon is clear. Regardless of what primary motivations exist for it, the choice of Order in Nêhiyawêwin is a gradient, rather than categorical, choice.

Bibliography

Abdulrahim, D. (2013). *A corpus study of basic motion verbs in Modern Standard Arabic*. PhD thesis, University of Alberta.

Agresti, A. (2013). *Categorical Data Analysis*. Wiley-Interscience.

Ahenakew, A. (2000). âh-âyîtaw isi ê-kî-kiskêyihtahkik maskihkiy/They Knew Both Sides of Medicine: Cree Tales of Curing and Cursing Told by Alice Ahenakew, edited by HC Wolfart and F Ahenakew. *Publications of the Algonquian Text Society. Winnipeg: University of Manitoba Press*.

Alcázar, A. and Saltarelli, M. (2014). *The Syntax of Imperative Clauses: A Performative Hypothesis*, pages 103–155. Cambridge Studies in Linguistics. Cambridge University Press.

Arppe, A. (2008). *Univariate, Bivariate, and Multivariate Methods in Corpus-Based Lexicography: A Study of Synonymy*. PhD thesis.

Arppe, A. (2009). Linguistic Choices vs. Probabilities–How Much and What can Linguistic Theory Explain? *The Fruits of Empirical Linguistics*, 1:1–24.

Arppe, A. (2013). *Polytomous: Polytomous Logistic Regression for Fixed and Mixed Effects*. R package version 0.1.6.

Arppe, A. and Järvikivi, J. (2007). Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. 3(2):131–159.

Arppe, A., Schmirler, K., Harrigan, A. G., and Wolvengrey, A. (2020). A Morphosyntactically Tagged Corpus for Plains Cree. In *Papers of the Forty-Ninth Algonquian Conference. Michigan State University Press*.

Baayen, R. H. (2012). Mixed-Effects Models. In *The Oxford Handbook of Laboratory Phonology*, pages 668–677. Oxford University Press.

Barth, D. and Kapatsinski, V. (2018). Evaluating Logistic Mixed-Effects Models of Corpus- Linguistic Data in Light of Lexical Diffusion. In *Mixed-Effects Regression Models in Linguistics*, pages 99–116. Springer.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bear, G., Fraser, M., Calliou, I., Wells, M., Lafond, A., and Longneck, R. (1992). *kôhkominawak otâcimowiniwâwa/Our Grandmothers' Lives: As Told in Their Own Words, edited by F Ahenakew and HC Wolfart*, volume 3. University of Regina Press.

Biber, D., Conrad, S., and Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Bloomfield, L. (1930). *Sacred Stories of the Sweet Grass Cree*. New York: AMS Press.

Bloomfield, L. (1946). Algonquian. In *Linguistic structures of Native America*, volume 6, pages 85–129. Viking Fund Publications in Anthropology, New York.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation*, pages 69–94. KNAW.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4):417–451.

Cook, C. (2014). *The Clause-Typing System of Plains Cree: Indexicality, Anaphoricity, and Contrast*. Oxford University Press, Oxford; New York.

Cook, C. and Muehlbauer, J. (2010). A Morpheme Index of Plains Cree.

Cook, C. E. (2008). *The syntax and semantics of clause-typing in Plains Cree*. PhD thesis, University of British Columbia.

Cruse, D. A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford linguistics. Oxford University Press.

Dacanay, D., Arppe, A., and Harrigan, A. (2021). Computational Analysis versus Human Intuition: A Critical Comparison of Vector Semantics with Manual Semantic Classification in the Context of Plains Cree. In *Proceedings of the 4th Workshop on the Use of Computational Methods in the Study of Endangered Languages*, volume 1, pages 33–43.

D'Agostino, R. B., Chase, W., and Belanger, A. (1988). The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations. *The American Statistician*, 42(3):198–202.

Dahlstrom, A. (2014). *Plains Cree Morphosyntax*. Routledge Library Editions: Linguistics. Routledge.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.

DiMarco, C., Hirst, G., and Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI Spring Symposium on Building Lexicons for Machine Translation*, volume 1, pages 114–121.

Divjak, D. (2010). *Structuring the Lexicon*. De Gruyter Mouton, Berlin, New York.

Divjak, D. and Arppe, A. (2013). Extracting prototypes from exemplars what can corpus data tell us about concept representation? *Cognitive linguistics*, 24(2):221–274.

Divjak, D. and Gries, S. T. (2006). Ways of Trying in Russian: Clustering Behavioral Profiles. *Corpus linguistics and linguistic theory*, 2(1):23–60.

Domencich, T. A. and McFadden, D. (1975). Urban Travel Demand–A Behavioral Analysis. Technical report.

Edmonds, P. and Hirst, G. (2002). Near-synonymy and Lexical Choice. *Computational linguistics*, 28(2):105–144.

Ethnologue (2016). Cree, Plains.

Firth, J. R. (1957). A Synopsis of Linguistic Theory 1930-55. In *Studies in Linguistic Analysis* (*special volume of the Philological Society*), volume 1952-59, pages 1–32. The Philological Society, Oxford.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Sage Publications, Inc.

Frank, E. and Kramer, S. (2004). Ensembles of Nested Dichotomies for Multi-Class Problems. In *Proceedings of the 21st International Conference on Machine Learning*.

Gelman, A. (2016). The problems with p-values are not just with p-values. *The American Statistician*, 70(10).

Goodman, L. A. and Kruskal, W. H. (1959). Measures of association for cross classifications. II: Further discussion and references. *Journal of the American Statistical Association*, 54(285):123–163.

Goodwin, W. W. (2002). *Greek Grammar*. Bristol Classical Press.

Gries, S. T. (2003). *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. A&C Black.

Gries, S. T. and Divjak, D. (2009). Behavioral profiles: a corpus-based approach to cognitive semantic analysis. *New directions in cognitive linguistics*, pages 57–75.

Gries, S. T. and Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective onalternations'. *International journal of corpus linguistics*, 9(1):97–129.

Hamp, B. and Feldweg, H. (1997). Germanet–A lexical-semantic Net for German. In *Automatic information extraction and building of lexical semantic resources for NLP applications*.

Han, W., Arppe, A., and Newman, J. (2013). Topic marking in a shanghainese corpus: from observation to prediction. *Corpus Linguistics and Linguistic Theory*, 0(0):1–29.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Mit Press.

Harrell, F. E. (2017). Regression Modeling Strategies. *Bios*, 330(2018):14.

Harrigan, A. and Arppe, A. (2021). Leveraging English Word Embeddings for Semi-Automatic Semantic Classification in Nêhiyawêwin (Plains Cree). In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 113–121. Association for Computational Linguistics.

Harrigan, A. G. and Arppe, A. (2015). oswêw êkwa ê-tipiskâk, mâka mâcîw âhpô ê-mâcît? Paper presented at the 47th Algonquian Conference, University of Manitoba, Winnipeg, Canada.

Harrigan, A. G. and Arppe, A. (2023). Leveraging Majority Language Resources for Plains Cree Semantic Classification. In *Papers of the Fifty-Second Algonquian Conference*, pages 129–146. Michigan State University Press.

Harrigan, A. G., Arppe, A., and Wolvengrey, A. (2018). Toward a Detailed Plains Cree VAI Paradigm. In *Papers of the Forty-Seventh Algonquian Conference*, pages 129–146. Michigan State University Press.

Harrigan, A. G., Schmirler, K., Arppe, A., Antonsen, L., Trosterud, T., and Wolvengrey, A. (2017). Learning from the computational modelling of plains cree verbs. *Morphology*, 27(4):565–598.

Harris, Z. S. (1954). Distributional Structure. *Word*, 10(2-3):146–162.

Henri, T. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76(1):103 – 154.

Hirst, G. (1997). Context as a spurious concept. *CoRR*, cmp-lg/9712003.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley, 2nd edition.

Inkpen, D. and Hirst, G. (2006). Building and Using a Lexical Knowledge Base of Near-Synonym Differences. *Computational Linguistics*, 32(2):223–262.

Jacques, G. and Antonov, A. (2014). Direct/inverse systems. *Language and Linguistics Compass*, 8(7):301–318.

Jolley, C. (1983). Algonquian person hierarchy: Morphosyntactic or semantic? *Algonquian Papers-Archive*, 14.

Kâ-Nîpitêhtêw, J. (1998). *Counselling Speeches of Jim Kâ-Nîpitêhtêw, edited by F Ahenakew and HC Wolfart*. University of Manitoba Press.

Kâ-Nîpitêhtêw, J. (2007). *The Counselling Speeches of Jim Kâ-Nîpitêhtêw*. Univ. of Manitoba Press.

Kassambara, A. and Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.

Klavan, J. (2012). *Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy*. PhD thesis, University of Tartu.

Klavan, J. (2020). Pitting corpus-based classification models against each other: a case study for predicting constructional choice in written estonian. *Corpus Linguistics and Linguistic Theory*, 16(2):363–391.

Klavan, J. and Divjak, D. (2016). The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica*, 50(2):355–384.

König, E. and Siemund, P. (2007). Speech act distinctions in grammar. *Language typology and syntactic description*, 1:276–324.

Lakoff, G. (1984). Performative subordinate clauses.

LeClaire, N., Cardinal, G., and Hunter, E. (1998). *Alberta Elders' Cree Dictionary/alperta ohci kehtehayak nehiyaw otwestamâkewasinahikan*. University of Alberta.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Lewis, P. M. and Simons, G. F. (2012). Assessing Endangerment: Expanding Fishman's GIDS. *Revue roumaine de linguistique. Editura Academiei: Bucuresti*, 55(2):103–120.

Majewska, O., Vulić, I., McCarthy, D., and Korhonen, A. (2020). Manual Clustering and Spatial Arrangement of Verbs for Multilingual Evaluation and Typology Analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4810–4824.

Masuskapoe, C. (2010). *piko kîkway ê-nakacihtât: kêkêk otâcimowina ê-nêhiyawastêki, edited by HC Wolfart and F Ahenakew*. Algonquian and Iroquoian Linguistics.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. pages 105–142. Academic Press.

McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. Technical report.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT press.

Minde, E. (1997). *kwayask ê-kî-pê-kiskinowâpahtihicik/Their Example Showed Me the Way, edited by F Ahenakew and HC Wolfart*. Edmonton: University of Alberta Press.

Okimāsis, J. L. (2018). *Cree: Language of the Plains/nēhiyawēwin: paskwāwi-pīkiskwēwin*. University of Regina Press.

Pearson, K. (1900). χ. On the Criterion That a Given System of Deviations From the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen From Random Sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175.

Pijpops, D. (2020). What is an Alternation?: Six Answers. *Belgian Journal of Linguistics*, 34(1):283–294.

Proulx, P. (1991). Proto-algic III. *Kansas Working Papers in Linguistics*, 16:129–170.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratt, S. (2016). *Mâci-nêhiyawêwin = Beginning Cree*. Regina : University of Regina Press.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Sadock, J. M. and Zwicky, A. M. (1985). *Speech Act Distinctions in Syntax.* Cambridge University Press.

Schmirler, K. (2022). *Syntactic Features and Text Types in 20th Century Plains Cree: A Constraint Grammar Approach*. PhD thesis, University of Alberta.

Schmirler, K., Arppe, A., Trosterud, T., and Antonsen, L. (2018). Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (*LREC 2018*), Miyazaki, Japan. European Language Resources Association (ELRA).

Searle, J. R. and Vanderveken, D. (1985). Speech acts and illocutionary logic. In *Logic, thought and action*, pages 109–132. Springer.

Snoek, C., Thunder, D., Loo, K., Arppe, A., Lachler, J., Moshagen, S., and Trosterud, T. (2014). Modeling the Noun Morphology of Plains Cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 34–42. Association for Computational Linguistics.

Statistics Canada (2016). Language Highlight Tables, 2016 Census.

Takahashi, H. (2008). Imperatives in concessive clauses: compatibility between constructions. *Constructions*.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76(1):103–154.

Tzougas, G. and Kutzkov, K. (2023). Enhancing logistic regression using neural networks for classification in actuarial learning. *Algorithms*, 16(2):99.

Vandall, P. and Douquette, J. (1987). *wâskahikaniwiyiniw-âcimowina/Stories of the House People, edited by F Ahenakew*. University of Manitoba Press.

Visitor, L., Junker, M.-O., and Neacappo, M. (2013). Eastern James Bay Cree Thematic Dictionary (Southern Dialect). *Chisasibi: Cree School Board*.

Wang, S. and Bond, F. (2013). Building the Chinese Open Wordnet (COW): Starting from Core Synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 10–18.

Whitecalf, S. (1993). kinêhiyâwiwininaw nêhiyawêwin/The Cree Language Is Our Identity: The La Ronge Lectures of Sarah Whitecalf, edited and translated by HC Wolfart and F Ahenakew. *Publications of the Algonquian Text Society / Collection de la Société d'édition des textes algonquiennes. Winnipeg: University of Manitoba Press*.

Wolfart, H. C. (1973). *Plains Cree: a Grammatical Study*. Transactions of the American Philosophical Society: new ser., v. 63, pt. 5. Philadelphia, American Philosophical Society, 1973.

Wolfart, H. C. (1996). Sketch of Cree, an Algonquian Language. In *Handbook of American Indians. Volume 17: Languages*, volume 17, pages 390–439. Smithsonian Institute, Washington.

Wolvengrey, A. (2001). *nêhiyawêwin itwêwina = Cree: Words*. University of Regina Press, Regina, bilingual edition edition.

Wolvengrey, A. (2005). Inversion and the Absence of Grammatical Relations in Plains Cree. *Morphosyntactic Expression in Functional Grammar*, 27.

Wolvengrey, A. (2011). *Semantic and Pragmatic Functions in Plains Cree Syntax*. PhD thesis, University of Amsterdam.

Wolvengrey, A. (2012). The Verbal Morphosyntax of Aspect-Tense-Modality in Dialects of Cree. Paper presented at the 2nd International Conference on Functional Discourse Grammar, Universiteit Ghent, Ghent, Belgium.

Yu, L.-C., Wang, J., Lai, K. R., and Zhang, X. (2017). Refining Word Embeddings for Sentiment Analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 534–539.