A Quantitative Account of Nêhiyawêwin Order: Using mixed-effects modelling 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

Abstract

This dissertation investigates the underpinnings of the phenomenon of Order in Nêhiyawêwin (Plains Cree) using quantitative methods and the Ahenakew-Wolfart Corpus (Arppe, Schmirler, Harrigan, & Wolvengrey, 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 dissertation 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 traditionally 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, as in Bresnan, Cueni, Nikitina, and Baayen (2007), Arppe (2008), and Divjak (2010). Specifically, mixed-effects logistics regression is used, allowing for modelling that takes into account the effects of sampling via random effects alongside fixed-effects. 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), morphosyntactic and semantic 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). An adjusted version of Chapter 2 has been accepted for publication in a future issue of Linguistics Vanguard and is undergoing final proofing as: Harrigan, A. G. & Arppe, A. (in press). Plains Cree Order as alternation. *Linguistics Vanguard*. This paper, which originated as Chapter 2, 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 4 were published as Harrigan, A. G., & Arppe, A. (2021). Leveraging English word embeddings for semi-automatic semantic classification in Nêhiyawêwin (Plains Cree). In M. Mager et al. (Eds.), *Proceedings of the first workshop on natural language processing for Indigenous languages of the Americas* (pp. 113–121). Association for Computational Linguistics. doi: https://doi.org/10.18653/V1/2021.AMERICASNLP-1.12; as well as Harrigan, A. G., & Arppe, A. (2023). Leveraging Majority Language Resources for Plains Cree Semantic Classification. In M. Macaulay M. Noodin (Eds.), Papers of the fifty-second Algonquian conference (pp. 129–146). Michigan State University Press. doi: https://doi.org/10.14321/j.ctv32r03jv.7. 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 evaluation criteria for the final product, and comparing purely automatic and semi-manual processes.

Acknowledgements

I would like to thank everyone who has helped me in the process of penning this dissertation. First and foremost is my supervisory committee. Antti Arppe has been a diligent and steadfast supervisor. Without his supervision, this work would not have been possible. Beyond academic and professional support, his assistance has helped shape my research ideology, my career trajectory, and my ability to remain resilient. Jordan Lachler has been a guiding force in my academic life since I began in Linguistics. Although I often felt like discussions with him left me less knowledgeable than prior, his approaches to language revitalization, morphological analysis, and e-mail etiquette remain a core influence to my own. Arok Wolvengrey and the expertise he brings in Nêhiyawêwin has been responsible for making this dissertation possible. Without the knowledge he has shared, I would simply not have been able to begin understanding this work. Finally, the expertise and advice offered by my external examiner, Graeme Hirst, has made the dissertation a much better piece of research.

I would like to express my gratitude to each of those speakers who participated in the Spoken Dictionary of Maskwacîs Cree: Mary Jean Littlechild, Louise Wildcat, Jerry Roasting, Harley Simon, Annette Lee, Arlene Makinaw, Rosie Roan, Rose Makinaw, Kîsikâw Kiseyiniw, Betty Simon, Brian Lightning, Brian Lee, Linda Oldpan, Mariam Buffalo, Debora Young, Ivy Raine, Agnes Fraynn, Paula Mackinaw, Norma Linda

Saddleback, Jack Mackinaw, Renee Makinaw, Linda Whitebear, Linda Wolfe, Ronald Littlechild, Lawrence Wildcat, and Marvin Littlechild. These speakers shared with me their language, culture, time, and friendship.

I would also like to thank my fellow students, particularly members of the Alberta Language Technology Laboratory. Among others, Katherine Schmirler and Eddie Antonio Santos have aided me socially, psychologically, and academically. Without these two, this dissertation would simply not have been possible. I also thank Megan Bontogon-Innes and Lex Giesbrecht for their collaboration throughout my academic career.

Without my friends and family, I would not have been able to finish this dissertation. I have relied on my parents, David and Tracy Harrigan in more ways than I can count. My friends have been similarly instrumental, and in particular I would like to thank the following for their unwavering support; without their guidance and acceptance, I can safely say I would not be where I am: McKinley Alden, Ashley Cipywnyk, Madeline Clark, Gregor Gronfeldt, Wayne Jackson, Matthew Kelley, Jared Millions, Keely Morrow, Thomas Saunders, and Kyle (Wally) Wicks.

This dissertation draws on research supported by the Social Sciences and Humanities Research Council.

Finally, I would like to thank the following individuals who, in addition to offering their support and care, have directly helped to edit this document: Zachary Berg, Taylor Casavant, Ella Madeline, Catherine Ford, Graham Mothersill, and Amanda Spencer. I owe these individuals a debt of gratitude that I cannot possibly pay back. Their expertise and friendship has made this document what it is.

Table of Contents

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
3″	Further 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

Corpus Code	Corpus File	Publication
CMBK-1-2	CMBK-CH1-VERS2	Masuskapoe (2010), Chapter 1
CMBK-2-2	CMBK-CH2-VERS2	Masuskapoe (2010), Chapter 2
CMBK-3-2	CMBK-CH3-VERS2	Masuskapoe (2010), Chapter 3
CMBK-4-2	CMBK-CH4-VERS2	Masuskapoe (2010), Chapter 4
CMBK-5-2	CMBK-CH5-VERS2	Masuskapoe (2010), Chapter 5
C2GB	C2GB-ARC	Bear et al. (1998), Chapter 2
AL	AL-RL-C.FIN	Bear et al. (1998), Chapter 2
SW	SW-CFIN2.D91	Whitecalf (1993)
VDC2	VDC2-RES.782	Vandall and Douquette (1987)
EM	EM-CREE5.N97	Minde (1997)
C7MW	C7MW-ARC.792	Bear et al. (1998), Chapter 7
JK	JK-C4ARC.798	Kâ-Nîpitêhtêw (1998)
C4MF	C4MF-ARC.792	Bear et al. (1998), Chapter 4
AA	AA-C2REV.899	Ahenakew (2000)
C6IC	C6IC-ARC.792	Bear et al. (1998), Chapter 6
C8GB	C8GB-ARC.792	Bear et al. (1998), Chapter 8

Chapter 1

Introduction

This dissertation explores the Nêhiyawêwin[1] phenomenon of Order. Nêhiyawêwin, like all Algonquian languages, is a polysynthetic language with a rich morphological system. Perhaps the most striking system in Algonquian verbal morphology is the system of Order.[2] 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.

Although Order is easily recognizable as a set of ways of inflecting verbs for person, it is a complicated phenomenon that is not easily applied to all verb stems in the same way, as will be described in Chapter 2. Further, the *purpose* of Order is unclear, despite attempts to describe the phenomenon 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 by Cook (2014, 11), who justifies her

[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.

[2]It is not clear why the term *Order* is used. The term appears to originate with Bloomfield (1946), though motivations for using this term are not well documented.

decision as thus: ‘There is a third paradigm: the imperative [O]rder. The imperative [O]rder 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 proposes that the phenomenon of Order is one of clause-typing, specifically in the difference between Indexical (not having a prior referent) and Anaphoric (having a prior referent) clauses. This conclusion was come to after careful hand-analysis of a Nêhiyawêwin corpus by Cook. Although I agree with many of the conclusions put forth by Cook, 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 that includes, in part, all of the texts used by Cook (2014) and modern computational techniques, this dissertation attempts to uncover why a speaker would choose to use one Order over another. This research is undertaken through the lens of alternation.

There are two research questions for this dissertation. They are presented below and numbered. The second research question is associated with specific predictions, enumerated below with alphabetical indices.

The main research question of this dissertation is as follows:

1. What morphosyntactic and semantic features affect a lemma’s propensity to occur in a particular alternation of Order?

Following from this main question, a secondary research question is presented. Here, two predictions are proposed:

2. Can Order choice be sufficiently predicted by primarily morphosyntactic and semantic predictors?

(a) Because Nêhiyawêwin is a morphologically rich language, and also due to the findings of previous similar alternation studies (Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013), the morphosyntactic and semantic predictors will provide substantial explanation of variation in modelling the alternations, though some variation will remain due to a lack of syntactic information.

(b) Semantic classification of constituents will do more to predict all alternations than morphosyntactic variables (as in Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013).

Adopting a usage-based approach based in the distributional hypothesis (Firth, 1962; 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. The primary method of analysis this dissertation relies on is mixed-effects logistic regression, based on and building upon the work of Bresnan et al. (2007), Arppe (2008), Divjak (2010), and 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, which is the most straightforward alternation in terms of the concept of near-synonymy (as used by Cruse, 2000, 157–159), and the alternation of the various Conjunct types (a more straightforwardly semantic alternation).

To lay the foundation for this research, 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 treatment of Order as an alternation, how the concept of alternation can be used to study the phenomenon, and a detailed justification for ignoring the Imperative mood (besides methodological opportunism).

Chapter 3 presents an overview of the corpus being used in this dissertation.

Next, Chapter 4 presents a study in the semi-automatically clustering of 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 effect in the statistical modelling of this dissertation.

Following the chapter of semantic classification, Chapter 5 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.

Chapter 6 presents the results of the statistical modelling in three stages: univariate, bivariate, and multivariate. As the last is of primary interest for this dissertation, it is presented most in-depth. The following chapter, Chapter 7, discusses in detail the multivariate results. This includes not only a discussion of what this means in the general sense of Order and 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 8 provides a conclusion to this dissertation.

Overall, this dissertation aims to provide a "linguistically informative" conception of Order. This can be defined as explanations or descriptions that appeal to linguistic motivations rather than settling for (significant) statistical association. For example, rather than focusing only on the fact that some individual feature combination, such as a verb in the past tense with a third-person actor, increases the likelihood of the Independent Order, this dissertation will attempt to provide an explanation of what this association might mean in terms of linguistic analysis and how results interact with previous descriptions of Order.

It is hoped that the results of this dissertation act as a basis for future, multi-disciplinary, and multi-method research. The results from this corpus-based research can and should be used to best select experimental stimuli to further investigate Order. Thus, a major motivation for this research is the production of hypotheses for further research.

Code used for the analyses presented in this dissertation is publicly available.[3] The underlying corpus that is analyzed is not able to shared publicly, however. 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.[4]

[3] https://github.com/atticusha/DissertationCode
[4] https://korp.altlab.app/

Chapter 2

Background

Nêhiyawêwin (ᓀᐦᐃᔭᐍᐏᐣ ISO 639-3: CRK, also known as Plains Cree) is a member of the Algonquian language family, which is one of the major language families of North America. Nêhiyawêwin is 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]

[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:

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*

Nêhiyawêwin is the westernmost member of the Cree-Montagnais-Naskapi continuum and is spoken mostly in Alberta, Saskatchewan, and northern Montana. Statistics Canada (2023) reports 12,005 individuals with knowledge of of 'Nehiyawewin (Plains Cree)' in Canada, though this number may be higher if respondents reported their language as 'Cree, not otherwise specified' rather than 'Nehiyawewin (Plains Cree).' Wolfart (1973) estimated 20,000 speakers, though the number has likely dropped since then. Although these numbers are 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 'Stable' by Ethnologue (Eberhard, Simons, & Fennig, 2023), which bases its classification on the Extended Graded Intergenerational Disruption Scale, a system for assessing language vitality based on domains of use, intergenerational transmission, and other sociolinguistic factors (Lewis & 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

[2] `https:/itwewina.altlab.app`

animate,[3] and thus the system is considered one of grammatical classification. Notably there are no, at least as far as I am aware, examples of animals that are grammatically inanimate. Animacy is relevant to nominal and verbal morphology in Nêhiyawêwin in various ways. Among nouns, this animacy distinction is manifested in two distinct plural markers, {-ak} for animate and {-a} for inanimate nouns; archaic/no-longer productive 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.

Inalienable nouns are more commonly referred to as *dependent nouns* in Algonquian linguistics and exist for both animate and inanimate classes. Dependent inanimate nouns are abbreviated *NDI* while dependent animate nouns are abbreviated *NDA*.[4] In general, NDIs are words representing body parts or some pieces of clothing, while NDAs mostly represented body parts and people of close relationships (e.g., *nikawiy*, 'my mother') (Okimāsis, 2018, 258). Dependent nouns differ from other nouns in that they require some level of possession to be uttered (Okimāsis, 2018, 257). For example, although

[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.

[4]As with the verbal inflectional classes, these abbreviations do not follow the expected English word order, in which case one might expect *NDI* and *NDA* as abbreviations. I am not clear on why this is the case, but it has become convention which I will continue to follow for sake of readability by Algonquianists.

the non-dependent/alienable inanimate noun *maskasin*, ‘shoe,’ may exist on its own (referring to a shoe that does not and has never belonged to any person), it can also be possessed as in *nimaskasin*, ‘my shoe.’ Similarly, the non-dependent/alienable animate noun *minôs*, ‘cat,’ is possessed by the first-person in the same way, creating *niminôs*, ‘my cat.’ Here, first-person possession follows the pattern described above. Dependent nouns, however, cannot occur without morphological possession. The word for ‘my heart’ is *nitêh*, but it is not grammatical to say *têh*. This requirement for a possessive morpheme is the same for animate and inanimate dependent nouns (Okimāsis, 2018, 257–271). If one wants to refer to a heart that is not possessed by a specific entity, the {m(i)-} prefix can be appended as in *mitêh*, ‘someone’s heart.’ These unspecified possessor forms are often those given as citation forms in dictionaries when listing dependent nouns. Note that alienable/non-dependent nouns cannot take this {mi-} suffix. There are some NDAs that do not allow for this unspecified possession, such as *-ôhkom*, the bare stem meaning ‘grandmother,’ which must be possessed by a first, second, or third-person to be used grammatically.

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 (a 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 ‘fourth-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 ‘fifth person,’ also with no number distinction; moving forward in this dissertation, it will be indicated by FUROBV). As obviation is based in topicality rather than syntactic roles (Bloomfield, 1946, 94), 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

Nêhiyawêwin verbs are traditionally classified according to both their transitivity and the animacy of their arguments/participants. There are two classes of intransitive verbs: one which can occur with one inanimate participant, called Intransitive Inanimate Verbs (VII)[5] and one which can occur with one animate participant, called Animate Intransitive Verbs (VAI). The former includes impersonal verbs (such as weather terms) and stative verbs used attributively to describe Inanimate objects. The VAI category 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.'

[5]The abbreviations and the ordering of the letters used for these classes are the standards used by Algonquianists.

(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.[6] 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. Obviaton, together with the directionality system discussed below, allow for semantic roles to be determined through relationships between items rather than through simple case marking.

[6] *Subjects* and *objects* are conventionally called *actors* and *goals* in Algonquian literature (Bloomfield, 1946; Wolvengrey, 2011). *Actors* here refer to the doer 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.

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 & 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[7] persons, wherein obviation applies. Due to this hierarchy, first-person acting on second necessarily always occurs with inverse morphology. This is 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).

[7]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).

(6) 2 > 1 > Unspecified Actor > 3 > 3′ > 3″

With obviation marked on both nouns and verbs, sentences such as those in (7a) 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 (7b). 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).'[8]

(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 inflectional classes. Alongside extensive person and direction morphology, several other categories may also be expressed on verbs.[9] Preverbs attach to the verb between person-prefixes 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

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

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

other Conjunct preverbs including {ka-}/{ta-}[10], 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.

In addition to these preverbs, Nêhiyawêwin also exhibits reduplication in the pre-stem position. These reduplication morphemes are not generally considered preverbs by convention. There are two reduplication templates which copy the initial consonant (C) from the morpheme they precede: {Ca-} and {Câh-}. The former, known as light (or weak) reduplication, indicates an ongoing action, and the latter, heavy (or strong) reduplication, indicates a repeated action. Before vowels, the vowel is not copied but the morphemes surface as {ay-} and {âh-} respectively (though there appear to be some rare exceptions to this rule). Preverbs can be preceded by reduplication as well, according to the same template, and both forms of reduplication can occur sequentially (e.g., {Ca-Câh-}).

2.3 Nêhiyawêwin Order

Order is a notable feature of Algonquian languages. Despite the similarities in terminology, Order has nothing to do with word order (of the linear progression kind). Although a confusing term, it is the one cemented by Bloomfield (1946), though it is

[10]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.

unclear what the origin or motivation for this term was.[11] Empirically, Order refers to the phenomenon of there being multiple ways to encode the same polypersonal agreement-markers on a verb. Put another way, there are multiple exponents for the same person/number portmanteau morphemes. For example, the words *ninipân* and *ê-nipâyân* are both forms for the first-person singular form of {nipâ-}, 'sleep.' Traditionally, Algonquian languages have three Orders: the Independent, the Conjunct, and the Imperative.[12] These Orders are essentially the three 'bins' of polypersonal agreement in the language. Every verbal lexeme can take the Independent and the Conjunct Orders; however, Intransitive Inanimate Verbs cannot take Imperatives (grammatically it is possible to command only second-person actors, which VII cannot have).

From a purely morphological point of view, the three Orders can defined as so:

1. The Independent, where the VAI, VTI and VTA classes use the {ni-} prefix for first-person, the {ki-} prefix for second-person, no prefix for third-person, and a set of suffixes that mark both person and number.

2. The Conjunct, which takes one of the prefixes {ê-}, {ka-}/{ta-}, {kâ-}, or Initial Change[13], and then a specific set of person-marking suffixes that is different than those in the Independent.

3. The Imperative, which has no grammatical prefix (like the Conjunct) nor a person-prefix (like the Independent), but rather uses person-marking suffixes.

Other Algonquian languages have further divisions for each Order. These are sometimes called *modes*, though Nêhiyawêwin has lost these in the Independent. In many of ways, Order seems like a system of mood as instantiated in a language such

[11] Interestingly, despite being a term used in Algonquian linguistics with some frequency, I was unable to find any scholar who could provide a history of the term beyond Bloomfield (1946); however, Bloomfield's lack of justification for the term could suggest that it was already established by that point.

[12] Like Order, these names are somewhat confusing as the do not necessarily correlate to the general meaning of *Independent* or *Conjunct*, though they are perhaps more motivated than *Order*.

[13] Details of this and other morphs/processes are given later in this section.

as Latin, wherein partial alteration to the stem indicates different moods; however, in Nêhiyawêwin the alteration indicating different moods take place on the affixes. This is not entirely straightforward however: The Independent indicates a matrix verb; the Conjunct can indicate either matrix and embedded verbs; and the Imperative indicates the Imperative mood. Thus, the system of Order in Nêhiyawêwin is one where the alternating between sets of affixes can suggest (but does not always confirm) syntactic structure, or it can indicate a particular mood. Further, the 'alternating between sets of affixes' presents entirely different paradigm shapes (i.e., the Imperative Order is restricted to only forms including second-persons) and applicability (the Imperative is more restricted than the other Orders).

This understanding produces a non-cohesive phenomenon that serves different types of linguistic functions depending on the verb stem being used, the syntactic context surrounding it, and the persons it marks for. Because the boundaries of this phenomenon are ill-defined, the rest of this section will detail previous descriptions of Order in Nêhiyawêwin. Ultimately, I will suggest that these previous definitions are problematic so long as the Independent, Conjunct, and Imperative are considered equally. A new treatment of Order is motivated and proposed. I argue that Order can be analyzed as a set of alternations and 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 the three main types the previous literature has detailed. The Independent is composed 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 or obviative persons) and a set of suffixes for marking person (Bloomfield, 1946; Wolfart, 1973). The Conjunct is composed of forms that also mark for any argument and which take no person-prefixes along side one of a number of conjunct suffixes (though they do take any one of grammatical prefixes or the Initial Change (mentioned later in this section), indicating 'Conjunctness'). The Imperative, on the other hand, marks for only person-arguments involving the second-person, cannot be used without such an argument, 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 ternary system. This organizational scheme, 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 paradigm 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 present-simple (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 a prefix for the third and obviative persons. Independent forms are unable to take the {ê-} preverb (discussed in the next subsection) which has begun to function primarily as a marker of Conjunct constructions.

Table 2.1: VII Independent paradigm for {*mihkwâ-*}*, 'to be red.' Based on Wolvengrey* (*2011, 393*).

Stem	Theme	SAP Person	Obv	3SG	3PL	3′		Example
mihkwâ				w			3SG	mihkwâw
mihkwâ				w	a		3PL	mihkwâwa
mihkwâ			yi	w			3′SG	mihkwâyiw
mihkwâ			yi	w	a		3′PL	mihkwâyiwa

Table 2.1[14] describes the Independent VII paradigm. Notice that only third-person (and obviative) participants exist in this paradigm, and so no speech-act participant[15] (SAP) prefix 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. An excerpt of a VTA mixed subparadigm is shown in Table (2.2).[16] This subparadigm also contains

[14]Only those paradigms necessary for understanding the general shape and complexity of Order are presented in this dissertation. Full, detailed, paradigms are available in Wolvengrey (2011, 394–429)

[15]Speech-act participant here refers to first or second-persons.

[16]A local subparadigm excerpt is seen in Table 2.5.

Table 2.2: VTA Independent direct, mixed participant, paradigm excerpt for {wâpam}, ‘to see him.’ Based on Wolvengrey (2011, 401).

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′[17]		Example
ni	wâpam	â			w	ak		1→3PL	niwâpamâwak
ki	wâpam	â			w	ak		2SG→3PL	niwâpamâwak
ni	wâpam	â	nân			ak		1PL→3PL	niwâpamânân
ki	wâpam	â	naw			ak		21PL→3PL [18]	kiwâpamânawak
ki	wâpam	â	wâw			ak		2PL→3PL	kiwâpamâwak
	wâpam	ê			w			3SG→3′	wâpamêw
	wâpam	ê			w	ak		3PL→3′	wâpamêwak
	wâpam	ê		yi	w		a	3′→3″	wâpamêyiwa

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 similar, differing in their inclusion of a theme sign.[19]

Table 2.3: VAI Independent paradigm for {nipâ}, ‘to sleep.’ Based on Wolvengrey (2011, 395).

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′		Example
ni	nipâ		n					1SG	ninipân
ki	nipâ		n					2SG	kinipân
ni	nipâ		nân					1PL	ninipânân
ki	nipâ		(nâ)naw					21PL	kinipâ(nâ)naw
ki	nipâ		nâwâw					2PL	kinipânâwâw
	nipâ				w			3SG	nipâw
	nipâ				w	ak		3PL	nipâwak
	nipâ			yi	w		a	3′	nipâyiwa

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

[17] In these paradigms and the following paradigms in this section, the *Obviative* column and 3′ columns essentially work together as a circumfix.

[18] 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.

[19] *Theme* is used in this dissertation in a similar sense to how it is used in traditional Indo-European grammar (for example see Grundt, 1978). In the case of Algonquian, the theme sign is used to associate a stem with a particular paradigmatic shape.

Table 2.4: VTI Independent Paradigm for {wâpaht-}, 'to see it.' Based on Wolvengrey (2011, 398). Note the difference of theme sign for local and non-local participants.

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′	Example	
ni	wâpaht	ê	n					1SG	niwâpahtên
ki	wâpaht	ê	n					2SG	kiwâpahtên
ni	wâpaht	ê	nân					1PL	niwâpahtênân
ki	wâpaht	ê	naw					2PL	kiwâpahtênaw
ki	wâpaht	ê	wâw					2PL	kiwâpahtêwâw
	wâpaht	am			(w)			3SG	wâpahtam
	wâpaht	am			w	ak		3PL	wâpahtamwak
	wâpaht	am		(i)yi	w		a	3′	wâpahtamiyiwa

paradigm. This has led 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).

Table 2.5: VTA Independent direct, Local participant, paradigm excerpt for {wâpam}. Based on Wolvengrey (2011, 401).

Prefix	Stem	Theme	1SG/PL	2PL	Example	
ki	wâpam	i	n		2SG→1SG	kiwâpamin
ki	wâpam	i	nân		2SG/PL→1PL	kiwâpaminân
ki	wâpam	i		nâwâw[20]	2PL→1SG	kiwâpaminâwâw

Table 2.6: VTA Independent inverse, mixed participant, paradigm excerptfor {wâpam}. Based on Wolvengrey (2011, 401).

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′	Example	
ni	wâpam	ik(w)			w	ak		1←3PL	niwâpamikwak
ki	wâpam	ik(w)			w	ak		2←3PL	niwâpamikwak
ni	wâpam	iko	nân			ak		1PL←3PL	niwâpamikonânak
ki	wâpam	iko	naw			ak		21PL←3PL	niwâpamikonawak
ki	wâpam	iko	wâw			ak		2PL←3PL	niwâpamikowâwak
	wâpam	ik(w)			(w)			3SG←3′	wâpamik
	wâpam	ik(w)			w	ak		3PL←3′	wâpamikwak
	wâpam	iko		yi	w		a	3′←3″	wâpamikoyiwa

Tables 2.2 and 2.5 through 2.7 give a subset of an Independent VTA paradigm,[21] 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 (as discussed earlier) changes. Note that some dialects allow for third-person inverse forms with {-ikow} endings instead of {-ik}.[22] While the VTA Independent forms are decomposible, the Conjunct forms are not always so predictable.

Table 2.7: VTA Independent inverse, local participant, paradigm excerpt for {*wâpam*}*. Based on Wolvengrey* (*2011, 401*)*.*

Prefix	Stem	Theme	1SG/PL	2PL		Example
ki	wâpam	iti	n		2SG←1SG	Example
ki	wâpam	iti	nân		2SG/PL←1PL	Example
ki	wâpam	iti		nâwâw[23]	2PL←1SG	Example

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’ (IC), which is an Algonquian process where the first vowel in the verb stem (or sometimes verbal prefixes) is mutated; this can be thought of a system of very similar allostems where one allostem is used in the ‘Changed’ form and another in the ‘Unchanged’ form. According to Wolfart, those Conjunct verbs with both {-ih} and IC are Iterative and are named by him as such. Those without IC, but with {-ih}, impart conditionality and are what Wolfart terms the *Subjunctive*. Conjunct verbs with IC, but without {-ih}, are simply called *Changed* and are the most commonly used Conjunct form, though Wolfart (1973) notes that Initial

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

[22] Note that the {-iko-} morph derives from the {-ikw-} morpheme followed by an epenthetic /i/, the combination of which produces /iko/.

[22] Although this form appears to only mark for direct theme and second-person plural, it is interpreted as second-person plural acting on first-person singular.

[23] Although this form appears to only mark for inverse and second-person plural, it is interpreted as a first-person acting on a second-person plural.

Table 2.8: Wolfart's Conjunct modes. Based on Wolfart (1973, 45).

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

Change is beginning to fall out of use, being replaced instead by the use of the {ê-} preverb. 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 Conjunct verbs without IC or {-ih} are referred to as *simple* (Wolfart, 1973). A summary of this four-way distinction is found in Table 2.8. In this table, the + column indicates a form that occurs with IC, while the − column represents items that do not. Items in the + row represent items that end with the {-ih} morph, while those in the − row do not. In more-contemporary Nêhiyawêwin orthography, the {-ih} suffix 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 *Changed* (those with either Initial Change or an {ê-} preverb) and *Unchanged* (those without) modes (2014).[24] The Changed Conjunct is further split into three subtypes: the *Changed Conjunct*$_1$, the *Changed Conjunct*$_2$, and the *Iterative Changed Conjunct*.[25] Although three subtypes are titled *Changed* due to being historically derived from Changed forms, only the Iterative currently exhibits Initial Change. Changed$_1$ and

[24]Mode is often used in Algonquian terminology to refer to subdivisions of Order, but is often unrelated to modality.

[25]Where Wolfart (1973) identified an Iterative/Conditional morpheme as {-ih}, Cook (2014) follows the contemporary orthography.

Table 2.9: Cook's Conjunct modes. Based on Cook (2014, 125).

Submode	Subtype	Form	Gloss
Changed	Changed Conjunct$_1$	ê-apiyân	'I sit'
	Changed Conjunct$_2$	kâ-apiyân	'when I sit'
	Iterative	êpiyâni	'whenever I sit'
Unchanged	Irrealis Simple	ka-apiyân	'for me to sit'
	Subjunctive	apiyâni	'when/if I sit'

Changed$_2$ on the other hand, are marked with the {ê-} and {kâ-} preverbs respectively.[26] 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.[27]

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.

As with the Independent paradigm, the VII Conjunct paradigm marks only for the third and obviative persons, as in Table 2.10. Note that here and throughout, when a morpheme ends in /t/ and precedes another morpheme beginning with /i/, the /t/ affricates into a voiceless alveolar affricate (<c>).

Table 2.10: VII Conjunct paradigm for {mihkwâ-}. Based on (Wolvengrey, 2011, 393).

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′	Actor	Example
ê-	mihkwâ				k			3SG	ê-mihkwâk
ê-	mihkwâ				k	i		3PL	ê-mihkwâki
ê-	mihkwâ			yi	k			3′SG	ê-mihkwâyik
ê-	mihkwâ			yi	k	i		3′PL	ê-mihkwâyik

[26]Wolfart (1973) classifies these two types together as Changed Conjunct forms, deriving {kâ-} from {kî-}.

[27]Terminology for these terms varies among 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, Arppe, & Wolvengrey, 2018).

Table 2.11: *VAI Conjunct paradigm for {nipâ}. Based on (Wolvengrey, 2011, 395).*

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′	Actor	Example
ê	nipâ		yân					1SG	ê-nipâyân
ê	nipâ		yan					2SG	ê-nipâyan
ê	nipâ		yâhk					1PL	ê-nipâyâhk
ê	nipâ		yahk					21PL	ê-nipâyahk
ê	nipâ		yêk					2PL	ê-nipâyêk
ê	nipâ				t			3SG	ê-nipât
ê	nipâ				t	ik		3PL	ê-nipâcik
ê	nipâ			yi	t			3′	ê-nipâyit

Table 2.12: *VTI Independent paradigm for {wâpaht-}. Based on (Wolvengrey, 2011, 398).*

Prefix	Stem	Theme	SAP Person	Obv	3SG	3PL	3′	Actor	Example
ê	wâpaht	am	ân					1SG	ê-wâpahtamân
ê	wâpaht	am	an					2SG	ê-wâpahtaman
ê	wâpaht	am	âhk					1PL	ê-wâpahtamâhk
ê	wâpaht	am	ahk					21PL	ê-wâpahtamahk
ê	wâpaht	am	êk					2PL	ê-wâpahtamêk
ê	wâpaht	am			k			3SG	ê-wâpahtahk[28]
ê	wâpaht	am			k	ik		3PL	ê-wâpahtahkik
ê	wâpaht	am		(i)yi	t			3′	ê-wâpahtamiyit

Similar to the Independent, the Conjunct's VAI and VTI paradigms are strikingly similar. The main difference is the inclusion of an epenthetic /j/ in the SAP Person endings for the VAI paradigm (because all VAI stems end in vowels, and Cree disprefers vowel-vowel sequences), as well as the {-am} theme element in the VTI. These differences are exemplified in the differences between Tables 2.11 and 2.12. Note that in the 1PL ← 3PL and 21PL ← 3PL forms, when the SAP morpheme ending with /Cw/ precedes the 3PL morpheme beginning with /ɪC/, the two phonemes coalesce into /o/. This appears to be a general process (Wolfart, 1973, 80).

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

[28] In the VTI paradigm, /m/ before /k/ becomes /h/.

Table 2.13: VTA Conjunct direct, local participant, paradigm excerpt for {*mow*}*, 'to eat him.' Based on Wolvengrey* (*2011, 402*)*.*

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

Table 2.14: VTA Conjunct inverse, mixed participant, paradigm excerpt for {*mow*}*. Based on Wolvengrey* (*2011, 402*)*.*

Prefix	Verb Stem	Theme	Obv	SAP	3SG	3PL	Actor ← Goal	Example
ê-	mow			it		ik	1SG ← 3PL	ê-mowicik
ê-	mow			isk		ik	2SG ← 3PL	ê-mowiskik
ê-	mow	iko		yâhk		ik	1PL ← 3PL	ê-mowikoyâhkik
ê-	mow	iko		yahkw		ik	21PL ← 3PL	ê-mowikoyâhkok
ê-	mow	iko		yêkw		ik	2PL ← 3PL	ê-mowikoyêkok
ê-	mow	iko			t		3SG ← 3′	ê-mowikot
ê-	mow	iko			t	ik	3PL ← 3′	ê-mowikocik
ê-	mow	iko	yi		t		3′ ← 3″	ê-mowikoyit

'chunked' 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-inflectional class appendices of Wolvengrey (2011).

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 or request to do something immediately, while the Delayed Imperative refers to a command or request to do something later. Because the Imperative only encodes command or request forms, both the Immediate and the Delayed mark only for second-person forms. Consequently, VII inflectional class of verbs, which only encodes third-person and obviative actors, does not occur in the Imperative.

Table 2.15: VTA Conjunct inverse, local participant, paradigm excerpt for {mow}. Based on Wolvengrey (2011, 402).

Prefix	Verb Stem	Theme	1SG/PL	2PL	Actor ← Goal	Example
ê-	mow	it	ân		2SG ← 1SG	ê-mowiyan
ê-	mow	it		âhk	2SG/PL ← 1PL	ê-mowiyâhk
ê-	mow	it		akok	2PL ← 1SG	ê-mowiyêk

Table 2.16: VTA Conjunct direct, mixed participant, paradigm excerpt for {mow}. Based on Wolvengrey (2011, 419).

Prefix	Verb Stem	Theme	Obv	SAP	3SG	3PL	Actor → Goal	Example
ê-	mow			ak		ik	1SG → 3PL	ê-mowakik
ê-	mow			at		ik	2SG → 3PL	ê-mowakik
ê-	mow	â		yâhk		ik	1PL → 3PL	ê-mowâyâhk
ê-	mow	â		yahk(w)		ik	21PL → 3PL	ê-mowâyâhkok
ê-	mow	â		yêkw		ik	21PL → 3PL	ê-mowâyêkok
ê-	mow	â			t		2PL → 3PL	ê-mowât
ê-	mow	â			t	ik	3PL → 3′	ê-mowâcik
ê-	mow	â	yi		t		3′ → 3″	ê-mowâyit

Across the remaining three inflectional 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.

Table 2.17: VAI Imperative paradigm for {nipâ}. Based on (Wolvengrey, 2011, 395).

Verb Stem	Immediate	Delayed	Actor	Example
nipâ			2SG	nipâ
nipâ	tân		21PL	nipâtân
nipâ	k		2PL	nipâk
nipâ		hkan	2SG	nipâhkan
nipâ		hkahk	21PL	nipâhkahk
nipâ		hkêk	2PL	nipâhkêk

Table 2.18: VTI Imperative paradigm for {wâpaht-}. Based on (Wolvengrey, 2011, 398).

Verb Stem	Theme	Immediate	Delayed	Actor	Example
wâpaht	a			2SG	wâpahta
wâpaht	ê	tân		21PL	wâpahtêtân
wâpaht	amw	ik		2PL	wâpahtamok
wâpaht	amw		ihkan	2SG	wâpahtamohkan
wâpaht	amw		ihkahk	21PL	wâpahtamohkahk
wâpaht	amw		ihkêk	2PL	wâpahtamohkêk

Additionally, the second-person plural and all delayed forms contain an epenthetic /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 2.19: VTA Imperative, mixed participant, paradigm for {mow}. Based on (Wolvengrey, 2011, 403).

		Immediate		Delayed		Actor	Example
Stem	Theme	3SG	3PL	3SG	3PL		
mow		(i)	ik			2SG	mowik
mow	â	tân	ik			21PL	mowâtânik
mow		ihkw	ik			2PL	mowihkok
mow	â			hkan	ik	2SG	mowâhkanik
mow	â			hkahkw	ik	21PL	mowâhkahkok
mow	â			hkêkw	ik	2PL	mowâhkêkok

Table 2.20: VTA Imperative, local participant, paradigm for {mow}. Based on (Wolvengrey, 2011, 403).

		Immediate		Delayed		Actor	Example
Stem	Theme	1SG	1PL	1SG	1PL		
mow	i	n				2SG	mowik
mow	i		nân			21SG/PL	mowâtânik
mow	i	k				2PL	mowihkok
mow	i			hkan		2SG	mowâhkanik
mow	i				hkahk	21SG/PL	mowâhkahkok
mow	i			hkêk		2PL	mowâhkêkok

The Imperative paradigm 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.

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. It only marks for second-person actors 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 organizational scheme that place the Imperative separate from the Independent and Conjunct, which are more similar to each other. This is illustrated in Table 2.21.

Table 2.21: Description of Orders by mood.

Order	Mood Specific
Imperative	✓
Independent	✗
Conjunct	✗

As will be seen throughout the rest of this chapter, this pattern of two Orders being 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 differently 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 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 (2014) details the syntactic distribution of the Conjunct Order, explaining like Wolvengrey (2011), that the Conjunct can occur in subordinate (i.e.,dependent clauses). 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 2.22 (2014, 125).

Table 2.22: Description of Conjunct Orders. Based on 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 does not explicitly 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 & Zwicky, 1985, 174). Wolfart (1973) mentions that the imperative is often—but not exclusively—used alongside 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 conveys a statement rather than a command. 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, for example, *I'm staying because consider which girl pinched me* (Lakoff, 1984, 476), are almost categorically ungrammatical to my ear as a native English speaker. 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).

The second possibility is one where the Imperative occurs in both Matrix and Embedded clauses. In either of these situations, the syntactic system does not cleanly align with the morphological system of Order, as shown in Table 2.23.

Table 2.23: Description of Orders by embedding status.

Order	Matrix	Embedded
Imperative	✓	✓
Independent	✓	✗
$Conjunct_1$	✓	✓
$Conjunct_2$	✗	✓
Iterative	✗	✓
Simple	✗	✓
Subjunctive	✗	✓

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 phenomena. 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 system is similar to the morphological organization seen previously and is shown in Table 2.24.

Table 2.24: Description of Orders by sentence type.

Order	Imperative	Declarative	Interrogative
Imperative	✓	✓	✓
Independent	✗	✓	✓
Conjunct	✗	✓	✓

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).[29] 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 (...)
extremely PST-be.kind-3.PL DIST.PL FOC.PL person-PL (...)

kâ-kî-ohpikih-iko-yâhkik
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 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 describes anaphoric clauses as having *some* sort of semantic or syntactic relation with a licensor in another clause. 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-Iterative Subjunctive form is not included by Cook, and its placement remains unclear.

Focusing specifically on the Conjunct modes, Cook 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 does not.

[29] In this example, I have removed a false-start/hesitation point that was transcribed in the source material and replaced it with with '(...).' I have done this for simplicity, as this example only serves to demonstrate Cook's description of Order. The false-start/hesitation is not important for this purpose. Later, in the Chapter 6, I have included any false starts and hesitations when presenting examples for my results and analyses.

Like the Changed $Conjunct_1$ forms, simple Conjuncts are 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).[30]

Cook does not include the Imperative Order in her study, and it is difficult to determine where it would be placed in her organizational scheme. 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 clause 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, Alcázar and Saltarelli (2014, 111) describe the Imperative (regardless of any specific language) as 'encoding the (indexical) parameters of the speech act, such as participant roles, temporality and locality.' An adaptation of Cook's Order organizational scheme with the Imperative included is found in Table 2.25. Note that all rows below the Independent are forms of Conjunct.

Table 2.25: Description of Orders by clause type. Based on Cook (2014)

Order	Indexical	Presupposing	Veridical
Imperative	✓	✗	✗
Independent	✓	✗	✗
$Conjunct_1$	✗	✗	✓
$Conjunct_2$	✗	✓	✗
Iterative	✗	✓	✗
Simple	✗	✗	✗

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).

[30] 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.

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. To study Order, it is critical to operationalize what different modes exist. Rather than simply taking either the Wolfart (1973) or Cook (2014) approaches, I opt to use corpus/empirical evidence to define the Conjunct modes on a structural basis. In the subset of the Ahenakew-Wolfart corpus (Arppe et al., 2020) used for this dissertation (see Chapter 3 for more detail), there are a number of different Conjunct types, though not all of those seen above. In considering types of Conjunct in this corpus, 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. Thus, the corpus has the following Conjunct types:

- {ê-} Initial
- {ka-}/{ta-} Initial
- {kâ-} Initial
- Initial Change
- Subjunctive {-i}

Both Initial Change and Subjunctive forms have only a small number of tokens in the corpora available to me. Bare tokens with Conjunct endings, but lacking either the Subjunctive {-i} or IC, were excluded as speaker who I spoke with during fieldwork in Maskwacîs speakers considered them as ‘incorrect,’ though it is possible that these may be constructed as simply reduced speech in connected speech. Interestingly, there were no forms in the analyzed corpus that contained both a Subjunctive suffix *and* IC; that is the *Iterative* in Wolfart (1973) and Cook (2014). While the corpus lacked an Iterative, it did contain verbs with *only* IC,[31] a form seemingly missing in 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*.

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-person, third-person, or obviative actors.
- It cannot take person-marking prefixes.
- It does not concern syntactic clause-typing (and instead concerns speech-act level information).
- It does not occur in statements of conditionality.

Comparatively, the primary difference between the Independent and Conjunct Orders are the morphological exponents used in each Order. This results in a morphological system as visualised in Table 2.26. Using this table, one could minimally describe the Imperative as those forms that do not take non-second-person actors, the Conjunct as those forms which take second-person actors but do not take person-prefixes, and the Independent as those forms which do take person-prefixes. Although the corpus used in this dissertation does not include Iterative Conjuncts, one could include them as a type of

[31] 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.

Initial Change Conjunct (as both contain Initial Change). Alternatively, one could group the Iterative with the Subjunctive, thus creating a distinction between bare forms with an {-i} suffix and bare forms without.

Table 2.26: Description of Orders in this dissertation by exponence.

Order	Non-second Actor	Distinct Person-prefixes	Other Prefixes
Imperative	✗	✗	✗
Independent	✓	✓	✗
ê-Conjunct	✓	✗	✓
kâ-Conjunct	✓	✗	✓
ka-Conjunct	✓	✗	✓
IC	✓	✗	✗
Subjunctive	✓	✗	✗

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 by which Algonquian languages use various exponents to mark person on verbs. As stated from the outset, we can identify three Orders in this way:

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 and a specific set of person-marking suffixes that is different than those in the Order identified above (the Conjunct); and

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); and

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 Table 2.27, wherein Independent and Imperative are indexical, while the Conjunct is anaphoric. This places Conjunct apart from the Independent and Imperative Orders.

Table 2.27: Description of Orders by clause type. Based on Cook (*2014*). (*repeated from page 33*)

Order	Indexical	Presupposing	Veridical
Imperative	✓	✗	✗
Independent	✓	✗	✗
Conjunct_1	✗	✗	✓
Conjunct_2	✗	✓	✗
Iterative	✗	✓	✗
Simple	✗	✗	✗

Finally, if we consider Order purely in terms of semantics, we can define Order as a system of distinguishing speech acts (the imperative vs. the declarative and interrogative). In this classification, the Independent and Conjunct are not distinguished from one another by speech act. This is shown in Table 2.28.

Table 2.28: Description of Orders by sentence type. (*repeated from page 33*)

Order	Imperative	Declarative	Interrogative
Imperative	✓	✓	✓
Independent	✗	✓	✓
Conjunct	✗	✓	✓

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 are 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). 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) to produce a surface form as in equation (2.1).[32]

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

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 of Nêhiyawêwin grammar, as any claim about Order needs to be relevant to all three of these categories. For these reasons, I will present a reanalysis of Order 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 it is not Order, I propose that the Imperative is a construction that acts as an illocutionary-force indicating device (Searle & 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

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

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 the definitions of Pijpops (2020), 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 & 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*).[33]. 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 by Biber, Conrad, and Reppen (1998) and mentioned as a caution for taxonomic classification by DiMarco, Hirst, and Stede (1993). Framing a phenomena 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 paradigms remain constant.

Using the lens of alternation, I propose that Order can be studied with systematic quantitative methods, as put forth by Arppe (2008, 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

[33] Alternations as described by Levin (1993) would likely fall into the realm of syntactic-semantic alternations, as they functionally use syntactic structure to group verbs at the highest levels. This is exemplified in the fact that Section 2 is a set of alternations based on 'alternations involving arguments within the VP.' (Levin, 1993, 45)

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 motivates 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 to 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.[34] This differs from similar phenomena such as noun class. In this phenomenon, there are indeed alternating paradigms with similar or identical shapes differing in exponents but it is not the case that any noun can occur in any paradigm. Instead, the paradigm in which a lexeme occurs is functionally an attribute of a lexeme. Importantly, Order is only able to be discussed as a paradigmatic alternation when the Imperative is removed from the system, as the Imperative is not applicable to all verbal lexemes.

[34]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 (Forthcoming).

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 on 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? Adopting a usage-based approach based in the distributional hypothesis (Firth, 1962; Harris, 1954), this research will use 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.[35] 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.

[35]All conjunct forms other than ê-Conjunct and kâ-Conjunct are combined int a single class for this research. See Chapter 5 for an explanation as to why this is.

Chapter 3

Corpus

This section details the corpus used for this dissertation. It discusses the features of the corpus as well as the representativeness of the corpus.

3.1 The Corpus

The underlying corpus from which the data set used in this dissertation is the Ahenakew-Wolfart corpus (Arppe et al., 2020). This corpus was compiled by Arppe et al. (2020) primarily to serve as a basis for quantitative morphosyntactic analysis of Nêhiyawêwin.

The Ahenakew-Wolfart corpus is likely the largest morphosyntactically tagged corpus of all Canadian Indigenous languages (excluding Inuktitut 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 by Ahenakew, 2000; Masuskapoe, 2010; etc.), there is still a paucity 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 17 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 by Ahenakew (2000); Bear et al. (1998); Kâ-

Nîpitêhtêw (1998); Masuskapoe (2010); Minde (1997); Vandall and Douquette (1987), and Whitecalf (1993). These texts include personal accounts, formalized lectures, and more-informal conversations, along with retellings of sacred stories. 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 inflectional 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 (10) and (11).

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

‘S/he lived thus / made a living thus.’

(11) 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, Arppe, Trosterud, & Antonsen, 2018). Among other features, this parser marks tokens for their status as predicate, actor, or goal.

3.2 Corpus Representativeness

An important aspect of any corpus research is a clear description of the boundaries and limits of the corpus. This section will describe the corpus based on the recommendations of Egbert, Biber, and Gray (2022). A one-page report on the corpus based on the following section is available as Appendix A.

As previously mentioned, the corpus is made up of a number of different text types. According to Schmirler (2022, 122–123), there are six main types of text in the corpus: sacred stories, old-time stories, counseling stories, funny stories, and personal stories. These text-types can be considered registers and were compiled together without regard for their relative weight in terms of size.[1] The stories were collected, edited, and transcribed by H. C. Wolfart and Freda Ahenakew between 1970s and 1990s. Speakers varied from across the plains and differed in their levels of formal education. As such, the corpus represents a relatively narrow window of time and may not represent Nêhiyawêwin as spoken in today, particularly by young people. Given the continual process of language loss, it is possible that certain grammatical features of the language have begun to change or disappear. Finally, no evaluation has yet been done to determine the representativeness of the corpus.

Motivations for choosing these texts to make up the corpus are primarily driven by a form of convenience sampling. These texts are those that were already transcribed in a standardized way and were already edited with translations attached. Due to a pluricentric system of orthography (e.g., spellings of words may vary in unpredictable ways from community to community), this standardization is paramount to corpus usage. This is particularly important with regards to the parsing that took place, which would be significantly more difficult with non-standardized texts.

[1] An additional set of texts, the so-called Bloomfield (sub)corpus (Arppe et al., 2020), was withheld from inclusion in the Ahenakew-Wolfart Corpus (at the time of writing this dissertation) due to the texts not having yet been syntactically tagged.

Given the above facts, this corpus cannot be considered an ideal representation of written Nêhiyawêwin as it is spoken in the present day (2023). Indeed, it is likely not an ideal representation of the language as spoken between the 1970s and 1990s, when the texts underlying it were collected. The corpus is simply too small, sampled by convenience, non-proportionally stratified, and lacks a robust corpus evaluation (though it has been used successfully in the 2022 work of Schmirler). What the corpus does offer is a unique opportunity to investigate morphosyntactic phenomenon in a lesser-studied language of Canada through a relatively diverse number, gender, and geographical distribution of speakers.

The diversity in this corpus is a major boon, as it broadly covers several aspects of Nêhiyawêwin. One must be clear, however, that the results of research based on this corpus may be skewed and should be verified in other samples of speech from other speakers along with the inclusions of further strata (with other samples of text, should they become available). Further, because it is likely that any attempt in the near future to make use of corpora in Nêhiyawêwin will likely continue to be skewed simply due to the lack of available resources (e.g., fluent, literate speakers to record and perform standardization of orthography), researchers should consider other avenues of research such as phonetic or psycholinguistic experimentation. Although these still suffer due to lack of available speakers, they can be used to avoid issues of orthography.

For a corpus of this size, especially one of a traditionally marginalized language, this corpus is quite robust.

Chapter 4

Automatic Semantic Classification

Abstract

Previous versions of this chapter were published as Harrigan and Arppe (2021) and Harrigan and Arppe (2023). Small, non-substantive changes have been made as per the external examiner for this dissertation. Some explanatory footnotes have also been added. This chapter 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 post-processing. 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.

The Ahenakew-Wolfart corpus used for this chapter is an older version of the corpus described in the previous section. The earlier corpus had four fewer tokens when considering only verbs in the Independent and ê-Conjunct forms, due to the continual improvements in recognition that took place between this research and the research of the rest of the dissertation.

4.1 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, Wang, Lai, and Zhang (2017) use polarity classifications (*good* vs. *bad*) from pre-defined lexica such as WordNet (Fellbaum, 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 of 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 (Fellbaum, 1998), German (Hamp & Feldweg, 1997), and Chinese (Wang & Bond, 2013), 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, Junker, & Neacappo, 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 such as Rapidwords[2] or even Levin's-style 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, Arppe, and Harrigan (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/`

4.2 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 myself, 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 made use of Word2vec (Mikolov, Chen, Corrado, & Dean, 2013) 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.[3] There are a number of other popular techniques to group words or sentences together. One of the most popular, tf-idf, identifies the most common words in a document and can then be combined with a clustering algorithm to group documents with similarly important terms.

Another technique is Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998), abbreviated LSA, which uses the concept of semantic vector space to define documents by vectors of word co-occurrences. Because raw vector spaces are usually sparse, LSA

[3]Although other word embedding methods exist, especially the currently popular BERT (Devlin, Chang, Lee, & Toutanova, 2019), when the research for this chapter was originally conceived, produced, presented and accepted for publication as a conference proceedings paper, Word2vec was still a primary methodology. Regardless, there is justification in using Word2vec, as the system is well-studied, documented, and understood. It relies on relatively simple computational tools and a conceptually simple methodology that are likely to be familiar and intuitive to many field linguists, even those without significant computational background. Author familiarity with and understanding of the technique are also reasons Word2vec was chosen as opposed to other similar vector approaches like GloVe (Pennington, Socher, & Manning, 2014). That said, any tool that produces vector representations of language can be used with the methodology described in this chapter.

uses singular value decomposition to reduce the vector space to some predetermined number of dimensions. Documents which share similar vectors in this new space can then be grouped together, often by cosine distance (Landauer et al., 1998, 269).

Similarly, Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003), abbreviated LDA, takes a collection of documents and attempts to model each document as a collection of some pre-defined number of latent topics. Each topic is also modeled a distribution over words. Unlike LSA which uses singular value decomposition, LDA is a generative model. The result of LDA is a set of documents, each defined by their mixture of multiple topics. One can then choose to group documents together based on their topic mixtures.

These techniques all consider language in a bag-of-words fashion; that is, they consider words as unrelated units without an organization structure. An alternative to this, called Continuous-Bag-of-Words (CBOW), considers the context of word (based on some defined window) instead of dealing only with a word in isolation. There is also the Skip-Gram approach which uses knowledge of a word to predict that words context, essentially the inverse of CBOW. These final two approaches are used by Word2vec to create word embeddings. Because I aim to use whole sentences to cluster words (as is discussed below), the ability to take into account context was an important factor in my not choosing these previously mentioned bag-of-words techniques.

Unfortunately, the Word2vec method is sensitive to corpus size. I initially attempted to create basic word and feature value co-occurrence matrices based on a 140,000-token Nêhiyawêwin corpus (Arppe et al., 2020) to create word vectors using Principal Components Analysis (PCA), 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 pre-processing 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 that of 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.

4.3 Methodology

To address the issue of corpus size, I attempted to bootstrap my classification scheme with pre-trained English vectors trained on 100 billion words of the Google News Corpus, from which a 3-million embedding model is derived. Each embedding represents a word through a 300-dimensional vector.[4]

I made use of the English definitions (specifically, the English words, and often full sentences, used to define Nêhiyawêwin words; sometimes also referred to as *glosses*) provided in Wolvengrey (2001) and fit to each word its respective Google News Corpus model's vector. This dictionary makes use of lemmas as headwords, and contains, at the time of writing, 21,717 entries. The presumption is that, because of the way the dictionary was compiled, 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 model, synonyms were used (one such rare example might be the word *mitêwin*, which is given

[4] This model is available at `https://code.google.com/archive/p/word2vec/`

in the definition as if it is an English word, and is unavailable in the model and thus would be replaced with something like *medicine lodge* or deleted if a synonym was given in the definition as well. This sort of situation was rare). Because the Google News Corpus model 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 pre-processing is not performed, these words are simply unavailable for clustering, as they lack a matching vector.[5] Where a Nêhiyawêwin word had more than 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 English definitions were lemmatized.

Once every English definition of every Nêhiyawêwin word in the Wolvengrey (2001) entries matched an entry in the Google News Corpus model, I associated each word in an English definition with its respective Google News Vector. That is, given a definition for the word *awâsisihkânis* as ‘small doll,’ the resulting structure of vectors would be:

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

Because all word-vectors in the Google News Corpus model 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

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

that acts as a sort of naïve sentence vector for each of the English glosses/definitions:

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

Mikolov, Sutskever, Chen, Corrado, and Dean (2013) mention this sort of naïve representation and suggest the use of phrase vectors instead of word vectors to address the representation of non-compositional idioms; however, given the way the definitions from Wolvengrey (2001) are written (e.g., with few idiomatic or metaphorical constructions), and for reasons of computational simplicity, I opted to use the above naïve implementation in this chapter.

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, 2022) to perform Hierarchical Agglomerative Clustering (HAC)[6] using the Ward method (Ward, 1963). The choice of HAC and the Ward method was 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 grouping item with the lowest error sum of squares of deviation from the

[6]Other methods for clustering also exist. Two of the most popular cluster algorithms are k-means clustering and DBSCAN. K-means clustering is a centroid-based algorithm that randomly assigns some number of centroids to a space, assigns all data points to the nearest centroid, adjusts the centroid's position based on its new members, and repeats this process until there is no change in cluster membership (Wu, 2012, 7). The number of clusters must be specified before hand and data clustered by k-means clustering cannot have multi-group membership. DBSCAN is a density based clustering method which does not cluster around a centroid, but rather on neighborhood membership (Ester, Kriegel, Sander, & Xu, 1996). Put simply, DBSCAN groups items together if they share neighbors in a defined space. DBSCAN is well suited to identifying clusters of arbitrary shape (Ester et al., 1996). Other methods of clustering such as Gaussian Mixture Model clustering, which is a method of clustering based on the modelling of distributions rather that centroid distance or density, have been applied to linguistic analysis (Li & Sporleder, 2010), but a complete overview of clustering options is beyond the purview of this chapter.

centroid of the cluster (Majerova & Nevima, 2017). This is an iterative process (Gries & Stoll, 2009) that starts pairing with individual data points until a single class of all points is created.

One of the main advantages of this method of clustering is that it 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 re-run the clustering. This method is very similar to what has been done by Arppe (2008), Bresnan et al. (2007), and Divjak and Gries (2006). For matters of familiarity and comparability, HAC was chosen. The choice of how many clusters were used (and thus where to cut the tree) was based on an impressionistic overview of effectiveness by myself.

For my purpose, I focused on the semantic classification of Nêhiyawêwin nouns and verbs. For the VIIs, 10 classes proved optimal, VAIs had 25 classes, VTIs with 15 classes, and VTAs with 20 classes. pre-processing verbs into these four classes was done, 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 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.[7] 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

[7] 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.

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 nouns 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.[8]

4.4 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. post-processing for these clusters took more time than other classes and essentially consisted 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 slightly fewer clusters after post-processing, though for VAIs this number was significantly higher, and for the NDIs it was slightly lower. Table 4.1 lists the number of clusters directly from HAC and from post-processing. 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 post-processing. For example, nearly all $\texttt{NI}_{14}$ items referred to parts of human bodies (and

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

Table 4.1: Hierarchical Agglomerative Clustering built cluster counts vs. counts after post-processing.

	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

those that did not fit this description were terms clearly related to, or containing, body parts like *aspatâskwahpisowin*, 'back rest')[9] while $\texttt{NI}_{13}$ was made up of trapping/hunting words and words for nests/animals.

The NA classes produced through HAC were similarly straightforward: $\texttt{NA}_9$ was made up of words for trees, poles, sticks, and plants; $\texttt{NA}_8$ was made up entirely of words relating to beasts of burden, carts, wheels, etc.; while much of $\texttt{NA}_3$ and $\texttt{NA}_7$, and nearly all of $\texttt{NA}_2$ referred to other animals. Once manually post-processed, the NA lexemes settled into eight classes: `NA-persons`, `NA-beast-of-burden`, `NA-food`, `NA-celestial`, `NA-body-part`, `NA-religion`, `NA-money.count`, and `NA-shield`.[10]

The NDI and NDA classes required almost no post-processing: $\texttt{NDA}_1$ and $\texttt{NDA}_3$ were each made up of various family and non-family-based relationships, while $\texttt{NDA}_2$ was 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`.

[9] While Nêhiyawêwin body parts are often dependent, they are not always so.

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

Verbs, on the other hand, required more post-processing. VIIs showed the best clustering results without post-processing. For example, $\texttt{VII}_6$ was entirely made up of taste/smell lexemes, $\texttt{VII}_7$ verbs were almost entirely weather-related, $\texttt{VII}_8$ contained verbs that only take plural subjects (the semantic nature of which is discussed below), $\texttt{VII}_9$ had only lexemes referring to sound and sight, and $\texttt{VII}_{10}$ had only nominal-like verbs (e.g., *mîsiyâpiskâw* '(it is) rust(y)').[11]

Despite these well-formed clusters, $\texttt{VII}_1$ through $\texttt{VII}_5$ were less cohesive and required manual clustering. In the end, six distinct classes were identified: `VII-natural-land`, `VII-weather`[12], `VII-sensory`, `VII-collective`,[13] `VII-move`, and `VII-named`.[14] Although post-processing was required, this was not too substantial in scope or time. The VAIs required significantly more work. Some classes were well-defined, such as $\texttt{VAI}_{23}$ whose members all described some sort of flight, but $\texttt{VAI}_{12}$ 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: `VAI-state`, `VAI-action`, `VAI-reflexive`, `VAI-cooking`, `VAI-speech`, `VAI-collective`, `VAI-care`, `VAI-heat/fire`, `VAI-money.count`, `VAI-pray`, `VAI-childcare`, `VAI-canine`[15], and `VAI-cover`. The VTIs similarly required manual post-processing after HAC clustering. Although some classes such as $\texttt{VTI}_{11}$ (entirely to do with cutting or breaking) or $\texttt{VTI}_{14}$ (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 six classes: `VTI-action`, `VTI-nonaction`,

[11] 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.

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

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

[14] This class contains terms of being named, such as *isiyihkâcikâtêw*, 'it is named thus.'

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

`VTI-speech`, `VTI-money.count`, `VTI-fit`, and `VTI-food`. Finally, the VTAs required a similar amount of post-processing as the VAIs. Although a few classes were well-formed (such as $\texttt{VTA}_4$ 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 post-processing, the following set of classes were defined: `VTA-action`, `VTA-nonaction`, `VTA-speech`, `VTA-food`, `VTA-money.count`, `VTA-religion`, and `VTA-allow`.

4.4.1 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 made use of the mixed-effects modelling that initially motivated this automatic semantic clustering, focusing on Nêhiyawêwin Order, specifically the alternation between the Independent and the ê-Conjunct, and using R (R Core Team, 2022) and the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015), I ran a mixed-effects logistic regression to predict alternation, with no fixed-effects but with semantic class as the sole random effect. So as to isolate the effect of semantic class, no other effects were used. Functionally, this creates models that consider each word in a corpus and whether it is observed in the Independent or ê-Conjunct. Using specified predictors, in this case the manually adjusted and HAC-only classes, the model can predict whether a word will occur in the Independent or ê-Conjunct Order. Two sets of models were fit for each of the four inflectional classes: one set using classes produced directly from the Hierarchical Agglomerative Clustering, and another set using those manually adjusted. Because semantic classes were introduced as a random-effect, the models take into account any inherent variation that exists between each of the classes and predict the outcome based on the information provided by the type of semantic classification *as a* *whole*. Put another way, the models predict Order based on the use of semantic-class information as a whole, rather than based on a particular semantic class. See Chapter 5 for more information.

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. See Chapter 3 for more information about this corpus.

To assess the effectiveness of semantic class in this context, I assessed Mcfadden's pseudo-R^2 (ρ^2) value. Because the models being evaluated are logistic (that is, they consider a binary choice: Independent vs. ê-Conjunct), a true R^2 score assessing estimated variance is inapplicable. Instead, a so-called pseudo-R^2 likelihood 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 a pseudo-R^2 measure 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 & 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 ρ^2 (Domencich & McFadden, 1975) as reported by the `ModelStatistics` function (Arppe, 2013). Mcfadden's ρ^2 appears to have a stable, but non-linear, relationship with a general R^2, wherein ρ^2 values of .2, .3, and .4 are roughly equivalent to an R^2 of .3, .5, and .73 respectively (Domencich & McFadden, 1975, 124). As with other pseudo-R^2 measures, a ρ^2 of over .25 is indicative of a fairly well-fitting model. Further, Han, Arppe, and Newman (2013) suggest that, in their experience, ρ^2 likelihood scores of nearly .30 are indicative of very good models without risk of over-fitting. As a general rule, a ρ^2 of 0.20 to 0.40 represents a well-fitting model (McFadden, 1977).[16]

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

Table 4.2: Model fitting results for 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

This method of evaluation was chosen as this semi-automatic clustering was originally motivated by this prediction and classification task.[17] The resulting ρ^2 scores from this procedure represent the way in which automatic and semi-manual clusters can explain the Nêhiyawêwin Order alternation.

Table 4.2 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 represents a model that is better able to predict the Order a verb token was observed in. 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 greater efficacy of fully-automatic 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 failed 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 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

[17] Other methods, such as the *silhouette method* used later in this dissertation could also have been used to evaluate the clustering, but they were not known to me at the time.

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 chapter 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, Vulić, McCarthy, and Korhonen (2020) should be considered.[18]

Table 4.3: 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)			

Table 4.3 details each of the post-processed 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 `NA-shield`

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

Table 4.4: Manually adjusted verb classes.

VII (N)	VAI (N)	VTI (N)	VTA (N)
`VII-natural-land` (275)	`VAI-state` (2,083)	`VTI-action` (1,409)	`VTA-action` (1,013)
`VII-weather` (96)	`VAI-action` (1,982)	`VTI-nonaction` (293)	`VTA-nonaction` (574)
`VII-sensory` (90)	`VAI-reflexive` (542)	`VTI-speech` (80)	`VTA-speech` (103)
`VII-collective` (79)	`VAI-cooking` (172)	`VTI-money.count` (25)	`VTA-food` (54)
`VII-move` (38)	`VAI-speech` (131)	`VTI-fit` (10)	`VTA-money.count` (23)
`VII-named` (3)	`VAI-collective` (97)	`VTI-food` (8)	`VTA-religion` (9)
	`VAI-care` (81)		`VTA-allow` (5)
	`VAI-heat/fire` (55)		
	`VAI-money.count` (34)		
	`VAI-pray` (29)		
	`VAI-childcare` (17)		
	`VAI-canine` (16)		
	`VAI-cover` (15)		

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 4.4 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 transitive 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 `VAI-cooking` and `VTI-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.

Verbs in `VAI-action` have little in common with each other except that they are a form of volitional action, while `VAI-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 `VAI-childcare` could be subsumed under `VAI-care`, the former includes items like *kimotôsêw* 's/he bears an illegitimate child.' This is even more obvious in categories such as `VAI-collective` or `VAI-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. As a result, 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, `VAI-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 post-processing. 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 on a prebuilt ontology; however, the appeal of a bottom-up classification should not be overlooked.

4.5 Discussion

In general, the best clustering was seen in classes with fewer items. The VAI and NI lexemes required the most post-processing, 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: 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 (12). 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`. Although one might assume that more definition words create a more specific (and thus accurate) embedding of the Nêhiyawêwin word, this would not explain why the longer definitions of verbs seemed to cluster in a less useful way than nouns. One possibility is that the increased number of words is essentially increasing noise (perhaps some of the words introduced are divergent enough from the 'true' sense of the Nêhiyawêwin word that they create a more generalized embedding), though this is speculation. More research on this is needed to understand this behaviour.

Because of the naïve nature of the averaging of the vectors, it is possible that definitions with more words create a more-generalized vector, rather than a more-specific one that better pinpoints a particular meaning.

(12) *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.

4.6 Conclusion

This chapter describes an attempt at semi-automatically classifying Nêhiyawêwin verbs and nouns. Resulting clusters of Nêhiyawêwin words are freely available online. Although the technique worked better with nouns, which required very little manual adjustment, verbs required more dedicated attention. Further, while ρ^2 was motivated by the underlying research question that led to this technique, it may not have been the best technique to evaluate these results. Thus, the impact of these results should not be overstated. Despite this, the technique presented in this chapter 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 naïve definition vectors for Wolvengrey (2001) with the same for the English WordNet word senses; word senses whose vectors bear a strong correlation with the English 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 well-resourced language, it is not sensitive to the issue of paucity of data for lesser-resourced languages. It should be applicable to any context where a lesser-resourced language has a majority-language bilingual dictionary and where the more-resourced 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 inflectional 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. This sort of 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, such as through cosine distance (as in Dacanay et al., 2021), should be considered. More-sophisticated sentence/definition embeddings, such as those returned by BERT (Devlin et al., 2019) 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 model. The technique described in this paper presents a compromise of taking the language on its own terms, while leveraging the large data sets that exist for better-resourced languages.

Chapter 5

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, Nêhiyawêwin Order can 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 on those univariate, bivariate, and multivariate statistics described by Divjak (2010), Bresnan et al. (2007), 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 by 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 4).

The corpus has also been 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.

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

Table 5.1: 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

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 4. 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 entries such as (13) and (14).

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

(14) `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 morphosyntactic or semantic tag 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 (13), the data frame extract in Table 5.1 is produced.

For the sake of fitting Table 5.1 to the page, the majority of the columns are not shown, but every morpheme observed in (13) would have a variable value of `TRUE` for the token `pimâtisiw`, and all morphemes 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 morphemes as one unit (e.g. `1Sg`), the data frame used for analysis in this dissertation splits the variables 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 inflectional 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 a VII in the corpus, while Wolvengrey (2001) analyzes it as a VAI, and LeClaire, Cardinal, and Hunter (1998) provides 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 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 eight unique classes: `Discourse`,[3]

[2]This process entailed inspecting the definition of all preverbs and grouping together those morphs with high-level, abstract similarities. 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.

Table 5.2: Preverb class tokens and types.

	Types	Tokens
Discourse	4	277
Position	15	285
Qual	30	316
Quant	7	10
Time	18	4,720
Move	4	731
Start/Finish	5	229
Want/Can	4	195

`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 5.2 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 4,777 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 by Arppe (2008) and allows for the assessment of individual values of categorical variables through straightforward application of chi-square analyses and logistic regression to predict a multinomial[4] alternation, in this case: Order.

5.1 Modelling the Alternations

In this dissertation, I will evaluate a univariate analysis given the morphosyntactic and semantic variables 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-Conjunct, Initial Change Conjunct, and Subjunctive Conjuncts were combined

[4]The term *multinomial* is used here instead of the term *polytomous* as by Arppe (2008). I use this term, except when referring to the specific package by Arppe (2013), as the term makes specific reference to nominal (not ordered) data, which is the sort of data of primary interest to this dissertation.

Table 5.3: Order tokens and types.

	Types	Tokens
Independent	876	4,390
ê-Conjunct	1,480	6,378
kâ-Conjunct	600	1,696
Other-Conjunct	393	828
Subjunctive	75	100
Initial Change	18	21
ka-Conjunct	344	707
Total	3,349	13,294

into a single 'Other' class. These three forms were combined primarily due to each of them being relatively infrequent and secondarily because each of these forms is more semantically-motivated than the other two conjunct forms. This joining of classes was chosen as opposed to simply discarding the Subjunctive and Initial Conjunct forms to maximize the number of tokens that could be analyzed. This results in the Order alternations as seen in Table 5.3.

To gain a wholistic understanding of Nêhiyawêwin Order, this dissertation will investigate three 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, 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. Worth noting is the fact that this final, multinomial, alternation is made up of outcomes that are more semantically motivated. The features the corpus is annotated with are primarily restricted to lexical semantics and morphosyntanx (rather

Table 5.4: *AWIvC* statistics.

	Types	Tokens
Independent	876	4,390
Conjunct	1,722	8,802
Total	2,598	13,192

Table 5.5: *AWIvE* statistics.

	Types	Tokens
Independent	876	4,390
ê-Conjunct	1,480	6,378
Total	2,356	10,768

than detailed higher-level semantic/pragmatic information of the utterance) and so may not be as predictive for this alternation. Despite this, the Conjunct-Type alternation was still investigated for any underlying morphosyntactic motivations.

Three main data frames were used:

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

In Tables (5.4) through (5.6) are relevant counts for each of the three data frames.

Table 5.6: *AWCnj* statistics.

	Types	Tokens
ê-Conjunct	1480	6,378
kâ-Conjunct	600	1,696
Other-Conjunct	393	828
Total	2,473	8,902

5.2 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 by Pearson (1900) and refined over the last century to produce the modern day chi-square 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 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 (5.1), where P is the SPR, 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 (5.1) the denominator represents its standard error.

$$P = \frac{O - E}{\sqrt{E(1 - t_i)(1 - t_j)}} \tag{5.1}$$

This produces SPR values that can be interpreted based on their 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 less than five. 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, Chase, & Belanger, 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.

In building models for univariate analysis, all variables with a minimum `TRUE` occurrence of 10 were selected for a given inflectional 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.

5.3 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.

There are a number of measures of association that can be used for assessing the degree of the relationship/bivariance of two categorical variables, as is the case in the alternations being studied for this dissertation. Some of the most common measures include: Pearson's Contingency Coefficient C (Goodman & Kruskal, 1954, 739; Liebetrau, 1983, 13), Tschuprow's Contingency Coefficient T (Goodman & Kruskal, 1954, 739–740; Liebetrau, 1983, 14), and Cramér's V (Cramér, 1946, 282–283, 443–444; Goodman & Kruskal, 1954, 740; Liebetrau, 1983, 14–15). Each of these measures may apply to nominal association (that is, with non-interval based variables) and are ultimately based upon the χ^2 statistic. Consequently, these measures depend on the dimensions of the underlying contingency table used in calculating the χ^2 statistic. As a result of this dependency, these measures can only be compared with other measures that are based on contingency tables of the same dimensions. Thus, an alternation with three outcomes is not comparable with an alternation with two outcomes (as the underlying contingency tables used to calculate the χ^2 statistic would be 2x2 and 2x3, respectively) using any of these above measures. Given that one of the three alternations investigated in this dissertation has three outcomes, this lack of comparability is of concern, even without taking into account the lack of compatibility with other studies.

In such cases, Arppe (2008, 87) suggests the use of measures based on Proportionate Reduction of Error (PRE), which measure how much classification error can be reduced (Costner, 1965) or how much a variable's variation can be explained by knowledge of another variable's distribution (Arppe, 2008, 87; Kviz, 1981). For the purposes of measuring bivariance, I will use Theil's Uncertainty Coefficient, U (Theil, 1970). Theil's U calculates the reduction of entropy in predicting one item given the observation of another (i.e. it calculates the reduction in uncertainty in predicting one item given the observation of another), bound between 0 (representing no reduction in entropy) and 1 (representing complete reduction in entropy). Essentially, it provides a measure of "relative reduction in uncertainty about Y from getting to know X" (Särndal, 1974, 171). *Uncertainty* in the context of this coefficient refers to viewing this bivariance in terms of how *certain* you can be of the presence of an outcome given the observation of another. Thus, a large coefficient represents more certainty or a reduction in uncertainty. As per Arppe (2008, 170), a value of .50 or higher can be considered a strong association; as such, this will be the cut off at which two variables are considered too collinear to both be included in further modelling.

An alternative PRE measure, the Goodman-Kruskal Tau exists. The Goodman-Kruskal Tau, τ, is a measure of how the knowledge of one variable predicts another. The measure considers the overall distribution of one variable (Variable A) as well as the distribution of Variable A given the presence of the values of Variable B. It then calculates the extent to which the latter increases prediction accuracy for the probabilities of all instances of Variable A. Put another way, the Goodman-Kruskal Tau is based on the baseline predictions specified by overall or conditional proportions of the dependent variable values. This measure is asymmetric for all contingency tables with dimensions larger than 2x2, though when measuring the association between two features (that is, a 2x2 contingency table), it becomes symmetric (meaning one is unable to measure the direction of the effect), while Theil's U maintains asymmetry (Arppe, 2008, 170; see also Costner, 1965, 351). For this reason, in measuring bivariance, Theil's U will be used.[5] Bivariate analysis for this dissertation makes use of the `associations` function from the `polytomous` package (Arppe, 2013).[6]

[5]See the appendices of Arppe (2008) for a detailed comparison of these and other various measures for association and bivariance.

[6]Although other measures such as Mutual Information (Church & Hanks, 1990) and 'T-scores' may be used as measures of association by some researchers, PRE measures are specifically created to measure association given contingency tables. As Stubbs (1995) points out, 'T-scores' are particularly problematic given their origin in the t-test which is not appropriate for categorical variables. For these reasons, and

5.3.1 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 $<$ .05). Automatic and manual classes were tested separately, as there was a great deal of bivariance between automatic and manual semantic-class variables. Information on bivariate pairs for each alternation and inflectional class is reported in Chapter 6.

5.4 Multivariate Analysis

5.4.1 Introduction

Using the methodology of Arppe (2008), and following bivariate analysis, the resulting variable sets (detailed in Chapter 6) 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 (Agresti, 2013, 163).

Logistic regression (e.g. with a single independent variable) can be modelled with the equation in (5.2) (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).

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \tag{5.2}$$

because the Goodman-Kruskal τ and Theil's U are part of a set of association measures created by statisticians as in Goodman and Kruskal (1954), Liebetrau (1983), Cramér (1946), and Theil (1970) (among others), I have not made use of these statistics in this dissertation.

The value resulting from (5.2) represents 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 occurring 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 (5.2) and seen in (5.3) (Agresti, 2013, 163).

$$logit[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x \tag{5.3}$$

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 the 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, 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 model 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. With regards to regression for language data, mixed-effects models have now become the norm (Barth & 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 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` or `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. Inclusion of lemma as a random effect is motivated by previous research by Harrigan and Arppe (2015), who found that certain lemmas were more likely to be observed in either the Independent or Conjunct Order than random chance would imply.

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 particular value for a variable is chosen as the baseline, then what is measured by observation is the opposite; that is, whether or not the value was **not** observed. The programming language used in this dissertation to statistically analyze results, R (R Core Team, 2022), by default uses the `0/FALSE` level as a baseline, 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 value. 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 value *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 the slope of `actor.1` 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).

5.4.2 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 alternation above, however, there is 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 OVR comparisons, a model compares one class against all other possible classes (Frank & Kramer, 2004). In this way it can be thought of as comparing x against $\neg x$. 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, whichever 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 can be straightforwardly described: a group of outcomes being compared is split repeatedly into mutually exclusive dichotomies until unary classifications are created (Frank & Kramer, 2004). The probability for each leaf/terminal node in the tree is the product of the probability of each of the previous outcomes/decisions made along on the way to the leaf node.

Similarly, there are other, alternative binarization techniques, such as designating one of the multinomial outcomes as a (possibly arbitrary) baseline/default/prototypical category, against which the other outcomes are compared (e.g. as described by Fox (2016, 392-393)). Arppe (2008) presents a detailed overview of all the above techniques as they relate to linguistic analysis; while Frank and Kramer (2004) present 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 a maximum of three outcomes in a single alternation.

One could use methods such as classification algorithms like C4.5 (Quinlan, 1993) to achieve similar results in place of logistic regression. Because the regression models are being used to essentially predictively classify verbs into particular Orders, such trees are a natural fit. Further, most classification-tree algorithms are designed to work with non-binomial decisions. Perhaps the most significant advantage of using a classification-tree algorithm is that the results of classification are easily visualized (James, Witten, Hastie, & Tibshirani, 2013, 315). Given that part of what drives this research is a linguistically informative conception of Order, the ability to easily visualize the decisions that ultimately lead to a classification is a major benefit. Despite these advantages, there are still good reasons to use logistic regression for the research in this dissertation. One obvious reason is that, despite the advantage that classification-tree algorithms have when dealing with multinomial decisions, only one of the three alternations being studied is actually multinomial. Further, and most important, is the fact that most commonly used classification-tree algorithms are not designed to account for random effects in the same way a mixed-effects regression model is. Although this is not in and of itself a substantial problem for these algorithms, previous research by Harrigan and Arppe (2015) motivates the assumption that lemma identity as a random effect is important in predicting Order. Finally, the field of alternation studies (for example in Abdulrahim, 2013; Arppe, 2008; Bresnan et al., 2007; Divjak & Arppe, 2013; Klavan, 2012; among others) makes extensive use of logistic regression. For these reasons, regression is used in this dissertation in place of other methods such as classification-tree algorithms.

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 calculate 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 is not without controversy (Wasserstein & Lazar, 2016), this dissertation will still use *p* values to determine which effects are most pertinent in modelling the Order alternation.

5.5 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 inflectional class (not for the model in its entirety). I have chosen not to present the standardized F-score for the reasons highlighted by Hand and Christen (2017); that is, the giving of equal weight to precision and recall by default in the F-score is unmotivated and can obscure nuance, especially when one deviates substantially from the other.

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 likelihood 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 & 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 & McFadden, 1975) as reported by the `ModelStatistics` function (Arppe, 2013). Mcfadden's ρ^2 appears to have a stable, but non-linear, relationship with a general R^2, wherein ρ^2 values of .2, .3, and .4 are roughly equivalent to an R^2 of .3, .5, and .73, respectively (Domencich & McFadden, 1975, 124). As with other pseudo-R^2 measures, a ρ^2 of over .25 is indicative of a fairly well-fitting model. Further, Han et al. (2013) suggest that, in their experience, ρ^2 likelihood scores of nearly .30 are indicative of very good models without risk of over-fitting. As a general rule, a ρ^2 of 0.20 to 0.40 represents a well-fitting model (McFadden, 1977).

In addition to pseudo-R^2, one can evaluate model performance using other measures. Just as with univariate and bivariate associations, models can be evaluated using non-χ^2 based methodologies. As previously mentioned, the measures of statistical relationships based on χ^2 suffer from a lack of comparability across studies with different underlying contingency table dimensionality. This means, for example, that alternations with different numbers of outcomes cannot be directly compared. As a form of proportionate reduction of classification error, Arppe (2008, 133) presents τ_p (Klecka, 1980, 51; Menard, 2002, 33), which he calls $\tau_{classification}$ (and which I will adopt, along with τ_c for short, for this dissertation). This measure is based on the previously mentioned Goodman-Kruskal Tau (Goodman & Kruskal, 1954, 745–747), and according to Arppe (2008, 133) is analogous to it. Applied to this dissertation's modelling, it represents the model's ability to predict Order as an outcome based on predictors rather than simply by using the distribution of the outcome as a whole. The measure generally ranges from 0–1[7] with higher values representing better-than-baseline classification. The Goodman-Kruskal Tau, and thus τ_c, is a relative measure that considers improvements of a model in the available prediction space, regardless of the actual value of baseline classification. Thus, increases in classification are reflected in a high τ_c value even when baseline classification is already quite high. Unlike the pseudo-R^2 measure, there do not appear to be agreed upon interpretations of what constitutes a good τ_c coefficient. Instead, the measure is better used as a way of comparing two (or more) models (as the model with the highest coefficient shows the largest increase in classification over baseline).

Although the results of the logistic regression are given in log-odds, this is straightforwardly transformed to probability by reversing the process of (5.2). We can do this by exponentiation of the log-odds we derived, as in (5.4):

$$p = \frac{e^x}{1 + e^x} \tag{5.4}$$

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 a resulting outcome given a (set of) predictor(s).

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

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 baseline against which one can assess how much of an

[7] In the most extreme cases, values can be negative, indicating a model does worse than baseline classification.

effect morphological features have on the ability to predict the type of Conjunct (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, pseudoR^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.

5.5.1 Research Questions and Predictions

There are two research questions for this dissertation. They are presented below and numbered. The second research question is associated with specific predictions, enumerated below with alphabetical indices.

The main research question of this dissertation is as follows:

1. What morphosyntactic and semantic features affect a lemma's propensity to occur in a particular alternation of Order?

Following from this main question, a secondary research question is presented. Here, two predictions are proposed:

2. Can Order choice be sufficiently predicted by primarily morphosyntactic and semantic predictors?

   (a) Because Nêhiyawêwin is a morphologically rich language, and also due to the findings of previous similar alternation studies (Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013), the morphosyntactic and semantic predictors will provide substantial explanation of variance in modelling the alternations, though some variation will remain due to a lack of syntactic information (e.g., argument structure).

(b) Semantic classification of constituents will do more to predict all alternations than morphosyntactic variables (as in Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013).

Chapter 6

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). In this case, the value is reported to three decimal places. If the *p* value is less than .001, the result is reported as < .001. Some results *appear* non-significant (i.e., having a reported *p* value of .05) due to rounding; however, these values are below .05 prior to rounding for presentation.

6.1 Univariate Results

Three separate univariate analyses were undertaken, one for each of the alternations being studied. Generally speaking, these analyses 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 included in the univariate models followed the selection criteria described in Chapter 5, 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 4 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 contain overt arguments. Further, previous syntactic accounts have not suggested linear order of arguments to be a substantial 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]

The selected variables were used with the `nominal` R function, producing a set of χ^2 test results. Statistically significant results are presented in Tables 6.1 through 6.4. These tables depict the predictor names, the number of tokens for each predictor, the χ^2 test statistic (with one degree of freedom, df, for the Independent vs. Conjunct and Independent vs. ê-Conjunct alternations, and two degrees of freedom for the Conjunct-Type alternation) for the predictor, the p value statistic for the test, and the direction of association between predictor and outcome (i.e., a $+$ indicates a significant positive association, which in turn implies that a predictor's `TRUE` value occurs significantly more often with a particular alternation construction than would be expected by chance, a $-$ represents the opposite, and a 0 represents no significant association).

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

6.1.1 Independent vs. Conjunct

Intransitive Inanimate Verbs

Table 6.1: Univariate results for the Independent vs. Conjunct alternation: VIIs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
`PV.time`	216	21.57	< .001	−	+
`II.sense`	275	32.49	< .001	−	+
`NI.object.actor`	145	4.87	.03	+	−
`Pron.actor`	58	5.20	.02	+	−
`Dem.actor`	57	6.42	.01	+	−
`Med.actor`	24	4.83	.03	+	−

For the Independent vs. Conjunct alternation in the VII class, only preverbs of time and sensory verbs were positively associated with the Independent Order. All other effects such as inanimate object actors, as well as pronominal, demonstrative, and medial actors were positively associated with the Conjunct. This is reported in Table 6.1. Note that both demonstrative actors and medial actors are features associated with certain pronouns, and so they are essentially subtypes of `Pron.actor`. As well, because `Med.actor` represents medial demonstrative pronouns, all `Med.actor` forms are necessarily also demonstrative actors forms.

Intransitive Animate Verb

The Independent vs. Conjunct VAI results showed a number of significant effects, nearly all positively associated with the Conjunct as detailed in Table 6.2. 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 appeared that semantic-effects were associated with the Conjunct,

[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 7. 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.

Table 6.2: *Univariate results for the Independent vs. Conjunct alternation: VAIs.*

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
`AI.speech`	1,348	1,112.49	< .001	−	+
`Actor.3`	3,663	118.87	< .001	−	+
`PV.time`	2,017	36.03	< .001	+	−
`PV.move`	444	43.51	< .001	+	−
`PV.qual`	173	8.52	.004	+	−
`PV.startfin`	143	9.00	.003	+	−
`PV.discourse`	137	27.58	< .001	+	−
`PV.position`	121	7.76	.005	+	−
`AI.do`	2,142	165.65	< .001	+	−
`AI.state`	1,943	89.47	< .001	+	−
`AI.cooking`	282	19.98	< .001	+	−
`AI.reflexive`	277	31.23	< .001	+	−
`AI.health`	128	6.70	.010	+	−
`AI.pray`	63	8.31	.004	+	−
`Rdplw`	142	12.17	< .001	+	−
`NA.persons.actor`	740	20.18	< .001	+	−
`Sg.actor`	541	9.92	.002	+	−
`Pl.actor`	297	11.32	.001	+	−
`Pron.actor`	405	6.36	.01	+	−
`Dem.actor`	201	5.42	.02	+	−
`NA.beast.of.burden.actor`	59	5.42	.02	+	−
`NA.food.actor`	37	4.36	.04	+	−
`Actor.1`	1,836	4.03	.05	+	−
`Actor.2`	265	16.88	< .001	+	−
`Actor.4`	178	34.61	< .001	+	−

with the semantic classes of action, state, cooking, reflexive, health, and praying all having significant positive Conjunct-effects. Beyond these and the effect of weak/light reduplication, all other remaining 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

Like in the Independent vs. Conjunct VAI model, there were a number of VTI model effects that showed a significant association, mostly with the Conjunct, as seen in Table 6.3. The only Independent-associated effects (though they accounted for roughly one third of tokens) were verbs having to do with cognition, actors that had referents that were

Table 6.3: Univariate results for the Independent vs. Conjunct alternation: VTIs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
`NI.nominal.goal`	115	9.96	.002	+	−
`NI.natural.force.goal`	73	4.48	.03	+	−
`NI.place.goal`	42	10.38	.001	+	−
`Sg.goal`	791	13.55	< .001	+	−
`Pl.goal`	248	15.89	< .001	+	−
`D.goal`	64	6.67	.010	+	−
`NDI.body.goal`	55	6.47	.01	+	−
`Px3sg.goal`	43	5.26	.02	+	−
`Der.dim.goal`	30	7.38	.007	+	−
`Actor.3`	1,514	142.00	< .001	+	−

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 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 are all associated with the Conjunct. The only actor-based effect for the Conjunct was that of third-persons actors, the opposite of the VTA results for the Independent vs. Conjunct alternation.

Transitive Animate Verbs

The VTA Independent vs. Conjunct results followed a similar pattern as seen previously. Nearly all significant effects in Table 6.4 were positively associated with the Conjunct, though verbs having to do with speech, having a first-person actor, and having a third-person goal were all positively associated with the Independent. As in the VAI results, all significant 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 significant in the VTA results. Second-person and obviative goals, those possessed by plural third-persons, and goals

Table 6.4: Univariate results for the Independent vs. Conjunct alternation: VTAs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
TA.speech	1,124	140.34	< .001	−	+
Goal.3	1,510	36.30	< .001	−	+
Actor.1	1,070	98.25	< .001	−	+
PV.time	1,189	52.78	< .001	+	−
PV.move	174	22.36	< .001	+	−
PV.discourse	68	12.02	.001	+	−
PV.qual	62	3.94	.05	+	−
PV.position	47	6.20	.01	+	−
TA.cognitive	852	5.74	.02	+	−
TA.do	839	59.42	< .001	+	−
TA.food	96	21.56	< .001	+	−
TA.money.count	66	4.13	.04	+	−
Goal.4	704	43.27	< .001	+	−
Goal.2	177	9.48	.002	+	−
NA.persons.actor	199	5.86	.02	+	−
Px3pl.goal	20	4.59	.03	+	−
NA.persons.goal	396	10.21	.001	+	−
NDA.relations.actor	83	5.22	.02	+	−
Sg.actor	179	12.25	< .001	+	−
D.actor	83	5.22	.02	+	−
Actor.3	1,277	26.13	< .001	+	−
Actor.4	152	17.72	< .001	+	−

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.

6.1.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Table 6.5: Univariate results for the Independent vs. ê-Conjunct alternation: VIIs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
II.sense	255	10.81	.001	−	+
PV.time	186	12.29	< .001	−	+
NI.object.actor	126	10.53	.001	+	−
Sg.actor	158	4.57	.03	+	−
Pron.actor	48	7.23	.007	+	−
Dem.actor	47	8.59	.003	+	−
Med.actor	21	6.64	.010	+	−

In the Independent vs. ê-Conjunct alternation, VII results demonstrated a positive association between sensory verbs and preverbs of time with the Independent outcome. Inanimate objects, singular, pronoun, demonstrative, and medial actors were all positively associated with the ê-Conjunct. Table 6.5 shows these effects.

Intransitive Animate Verbs

Table 6.6: Univariate results for the Independent vs. ê-Conjunct alternation: VAIs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
`AI.speech`	1,240	925.86	< .001	−	+
`Actor.3`	3,109	88.41	< .001	−	+
`PV.time`	1,654	38.79	< .001	+	−
`PV.move`	346	39.34	< .001	+	−
`PV.qual`	147	11.36	.001	+	−
`PV.startfin`	123	12.38	< .001	+	−
`PV.discourse`	116	33.00	< .001	+	−
`PV.position`	109	12.86	< .001	+	−
`AI.do`	1,671	133.84	< .001	+	−
`AI.state`	1,578	91.04	< .001	+	−
`AI.cooking`	231	21.99	< .001	+	−
`AI.reflexive`	222	32.02	< .001	+	−
`AI.pray`	46	6.36	.01	+	−
`Rdplw`	124	17.10	< .001	+	−
`NA.persons.actor`	554	8.44	.004	+	−
`NA.beast.of.burden.actor`	46	4.93	.03	+	−
`Pl.actor`	228	7.62	.006	+	−
`Actor.1`	1,551	9.79	.002	+	−
`Actor.4`	138	33.74	< .001	+	−

The Independent vs. ê-Conjunct VAI results continued previous trends: verbs of speech with third-person actors associated with the Independent, but all other significant effects as described in Table 6.6 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 6.7: Univariate results for the Independent vs. ê-Conjunct alternation: VTIs.

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
TI.cognitive	1,008	152.52	< .001	−	+
NA.persons.actor	203	8.92	.003	−	+
Pers.actor	95	16.13	< .001	−	+
Pron.actor	129	13.14	< .001	−	+
Px1sg.goal	15	3.85	.05	−	+
Actor.2	181	58.80	< .001	−	+
Actor.1	1,043	80.68	< .001	−	+
TI.do	1,281	146.95	< .001	+	−
TI.money.count	17	4.17	.04	+	−
PV.discourse	55	17.18	< .001	+	−
NI.natural.force.goal	64	7.16	.007	+	−
NI.place.goal	31	9.97	.002	+	−
Der.dim.goal	20	5.85	.02	+	−
Pl.goal	205	20.06	< .001	+	−
Actor.3	1,184	132.23	< .001	+	−

The Independent vs. ê-Conjunct VTI results 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 with third-person actors.

Transitive Animate Verbs

The VTA results in the Independent vs. ê-Conjunct alternation mostly exhibited 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

Table 6.8: *Univariate results for the Independent vs. ê-Conjunct alternation: VTAs.*

	N	$\chi^2_{df=1}$	$p\ (\chi^2)$	CNJ	IND
TA.speech	905	155.53	< .001	−	+
Actor.1	892	84.26	< .001	−	+
Actor.2	84	22.84	< .001	−	+
Goal.3	1,185	52.70	< .001	−	+
PV.time	946	59.05	< .001	+	−
PV.move	123	16.61	< .001	+	−
PV.discourse	58	16.41	< .001	+	−
PV.qual	49	4.33	.04	+	−
PV.position	44	11.25	.001	+	−
PV.startfin	30	4.54	.03	+	−
TA.cognitive	692	9.13	.003	+	−
TA.do	650	57.05	< .001	+	−
TA.food	80	27.39	< .001	+	−
Sg.actor	138	11.84	.001	+	−
D.actor	68	6.84	.009	+	−
NDA.relations.actor	68	6.84	.009	+	−
Actor.3	1,060	48.02	< .001	+	−
Actor.4	109	13.59	< .001	+	−
Goal.4	579	59.52	< .001	+	−
NA.persons.goal	290	4.42	.04	+	−
Px3sg.goal	36	6.28	.01	+	−
Px3pl.goal	15	4.64	.03	+	−

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. These effects are described in Table 6.8.

6.1.3 Conjunct-Type

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 6.9: Univariate results for the Conjunct-Type alternation: VIIs.

	N	$\chi^2_{df=2}$	$p\ (\chi^2)$	Ê-CNJ	KÂ-CNJ	OTHER-CNJ
`II.sense`	171	42.38	< .001	+	−	0
`PV.time`	135	13.47	.001	+	−	0
`Sg.actor`	134	28.09	< .001	+	−	0
`NI.object.actor`	119	16.49	< .001	+	−	0
`II.natural.land`	145	46.84	< .001	−	+	0
`II.weather`	126	17.44	< .001	−	+	0

The Conjunct-Type VII results 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 actors, and object actors were positively associated with the ê-Conjunct and negatively associated with the kâ-Conjunct. Verbs representing nature/land and weather were the odd ones out, positively associating with kâ-Conjunct and negatively associating with the ê-Conjunct. The effects are shown in Table 6.9.

Intransitive Animate Verbs

The Conjunct-Type VAI results were more varied, as evident in Table 6.10. 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 were 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

Table 6.10: Univariate results for the Conjunct-Type alternation: VAIs.

	N	$\chi^2_{df=2}$	$p(\chi^2)$	Ê-CNJ	KÂ-CNJ	OTHER-CNJ
`PV.discourse`	120	6.16	.05	+	−	0
`PV.position`	95	9.98	.007	+	−	0
`Actor.1`	1,251	14.26	.001	+	−	0
`Actor.3`	2,222	17.39	< .001	+	0	−
`Rdplw`	114	7.34	.03	+	0	0
`Actor.2`	207	212.19	< .001	−	+	+
`AI.do`	1,649	10.71	.005	−	+	0
`NA.persons.actor`	545	37.19	< .001	−	+	0
`Sg.actor`	392	21.98	< .001	−	+	0
`Pron.actor`	292	23.53	< .001	−	+	0
`Prox.actor`	91	25.03	< .001	−	+	0
`Dem.actor`	149	30.52	< .001	−	+	−
`AI.health`	99	14.38	.001	−	0	+
`Pl.actor`	224	19.45	< .001	0	+	−
`Med.actor`	56	6.19	.05	0	+	0
`PV.time`	1,442	90.04	< .001	0	−	+
`PV.wantcan`	45	39.12	< .001	0	−	+
`AI.cooking`	222	7.25	.03	0	−	0
`PV.qual`	133	7.54	.02	0	0	−
`D.actor`	164	11.51	.003	0	0	−
`NDA.relations.actor`	164	11.51	.003	0	0	−
`Px1Sg.actor`	123	8.03	.02	0	0	−

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, which can also be 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

The Conjunct-Type VTI results 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,

Table 6.11: Univariate results for the Conjunct-Type alternation: VTIs.

	N	$\chi^2_{df=2}$	$p\ (\chi^2)$	Ê-CNJ	KÂ-CNJ	OTHER-CNJ
`PV.position`	45	12.87	.002	+	−	−
`Actor.1`	693	17.46	< .001	+	0	−
`PV.wantcan`	51	22.80	< .001	−	+	0
`TI.speech`	171	34.91	< .001	−	+	0
`NA.persons.actor`	162	14.26	.001	−	+	0
`Sg.goal`	586	51.80	< .001	−	+	+
`NI.object.goal`	512	32.80	< .001	−	+	0
`Dem.goal`	216	39.97	< .001	−	+	0
`Pron.goal`	216	39.97	< .001	−	+	0
`Prox.goal`	122	40.87	< .001	−	+	0
`NI.nominal.goal`	95	23.66	< .001	−	+	+
`Actor.2`	137	105.38	< .001	−	0	+
`Med.goal`	94	6.08	.05	−	0	0
`D.goal`	54	6.36	.04	−	0	+
`PV.time`	860	59.54	< .001	0	−	+
`Sg.actor`	93	9.31	.010	0	+	0

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 preverbs. 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. 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. These effects are presented in Table 6.11.

Transitive Animate Verbs

Results for the VTAs present a number of different significant effects for all outcomes, detailed in Table 6.12. Preverbs of time, position, and discourse; third-person actors, first-person and obviative goals, and verbs of cognition were all positively associated with

Table 6.12: Univariate results for the Conjunct-Type alternation: VTAs.

	N	$\chi^2_{df=2}$	$p\ (\chi^2)$	Ê-CNJ	KÂ-CNJ	OTHER-CNJ
`PV.time`	863	59.58	< .001	+	−	+
`TA.cognitive`	580	19.78	< .001	+	−	+
`Actor.3`	893	35.56	< .001	+	−	−
`Goal.obv`	529	18.82	< .001	+	−	−
`PV.position`	39	10.17	.006	+	−	0
`Goal.1`	392	17.79	< .001	+	0	−
`PV.discourse`	58	6.93	.03	+	0	−
`TA.speech`	576	22.35	< .001	−	+	0
`Sg.goal`	195	16.31	< .001	−	+	0
`Prox.goal`	66	7.99	.02	−	+	0
`Actor.2`	84	121.61	< .001	−	0	+
`Goal.3`	897	25.28	< .001	−	0	+
`Goal.2`	134	10.29	.006	−	0	+
`Prox.actor`	31	9.99	.007	0	+	0
`Px1sg.goal`	67	6.44	.04	0	0	−
`Actor.1`	567	7.34	.03	0	0	−

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 were 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.

6.2 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 (1970) uncertainty coefficients of greater than .50, as in Arppe (2008), indicating that knowing the value of one variable provides substantial information about the value of another) were dealt with by removing one of the items in the pair. This is done as, according to Harrell (2015, 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 potential explanatory value as well as its frequency. For example, in some cases, pairs of variables like `D.goal` (dependent goal) and `NDI.body.goal` (dependent inanimate noun that represents a body part and acts as a goal) were bivariate. Every word that is tagged `NDI.body.goal` would also be tagged `D.goal` (as all `NDI.body.goal` are dependent). In a case such as this, `NDI.body.goal` was retained as it contained more semantic information and thus had a higher potential explanatory value. If each member of a bivariate pair was deemed to be of the same relevance to the modelling and had the same potential explanatory value, then the more frequent item was retained.

The following section presents the effects that formed bivariate pairs along with the list of effects to removed from consideration for modelling to resolve bivariance. The list of remaining 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* indicates the extent to which *Category1*'s value predicts the presence or absence of *Category2*'s value and *uc.21* gives the inverse).

6.2.1 Independent vs. Conjunct

Intransitive Inanimate Verbs

Table 6.13: 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	.29	.53
`NI.object.actor`	`Dem.actor`	144	57	57	.29	.53
`Pron.actor`	`Dem.actor`	58	57	57	.96	.98
`Pron.actor`	`Med.actor`	58	24	24	.33	.63
`Dem.actor`	`Med.actor`	57	24	24	.34	.64

Bivariance among 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 6.13 also show that nearly all pronominal actors were demonstrative and that all demonstrative pronouns were medial; less interestingly, all medial pronouns co-ocurred 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`, and `Med.actor`.

Intransitive Animate Verbs

Table 6.14: 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	.44	.66
`NA.persons.actor`	`Dem.actor`	737	201	201	.20	.51
`Pron.actor`	`Dem.actor`	403	201	201	.41	.69
`Actor.3`	`Actor.1`	3,646	1,836	0	.49	.56

The VAIs' bivariance for the Independent vs. Conjunct alternation focused on actor variables, shown in Table 6.14. 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 co-occurred with the former. I removed `Pron.actor`, `Dem.actor`, and `Actor.1` to resolve the 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`, and `Actor.4`.

Transitive Inanimate Verbs

Table 6.15: Bivariate results for the Independent vs. Conjunct alternation: VTIs.

category1	category2	N1	N2	N12	uc.12	uc.21
TI.do	TI.cognitive	1,632	1,163	0	.64	.66
NA.persons.actor	Pron.actor	261	158	158	.50	.72
NA.persons.actor	Pers.actor	261	107	107	.32	.62
Pron.actor	Pers.actor	158	107	107	.59	.79
Actor.3	Actor.1	1,508	1,202	0	.58	.60
D.goal	NDI.body.goal	64	55	55	.80	.91

The VTIs show a similar pattern as the above classes. Pronouns, and specifically personal pronouns, always occurred 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 co-occurred. Finally, the two main verb classes, `TI.do` and `TI.cognitive`, were also bivariate, never occurring together. These relationships are depicted in Table 6.15. To address the presence of 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`, and `Px1sg.goal`.

Transitive Animate Verbs

Table 6.16: 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	1,266	1,498	0	.68	.67

Table 6.16 shows that the VTA class 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). I removed `D.actor` and `Actor.3`, 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`, and `Px3pl.goal`.

6.2.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Table 6.17: Bivariate results for the Independent vs. ê-Conjunct alternation: VIIs.

category1	category2	N1	N2	N12	uc.12	uc.21
Sg.actor	NI.object.actor	158	126	119	.53	.60
NI.object.actor	Pron.actor	126	48	48	.27	.50
Pron.actor	Dem.actor	48	47	47	.96	.97
Pron.actor	Med.actor	48	21	21	.35	.64
Dem.actor	Med.actor	47	21	21	.36	.65

As seen in Table 6.17, the VII bivariate 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 occurred with Pron.actor. Similarly, Med.actor always occurred with both Pron.actor and Dem.actor. Removing Sg.actor, Pron.actor, and Med.actor left 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 6.18: Bivariate results for the Independent vs. ê-Conjunct alternation: VTIs.

category1	category2	N1	N2	N12	uc.12	uc.21
TI.do	TI.cognitive	1,281	1,008	0	.68	.69
NA.persons.actor	Pron.actor	203	129	129	.53	.74
NA.persons.actor	Pers.actor	203	95	95	.37	.65
Pron.actor	Pers.actor	129	95	95	.65	.82
Actor.3	Actor.1	1,184	1,043	0	.62	.63

As in the previous alternation, VTIs of cognition and verbs of action were bivariate in that they never occured together. Similar to other classes, Pron.actor and Pers.actor always occurred with NA.persons.actor, and Pers.actor did the same with Pron.actor. Finally, third and first-person actors never co-occurred. These relationships are shown in Table 6.18. To alleviate this covariance, the 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, and Px1sg.goal as the final set of variables for multivariate analysis.

Transitive Animate Verbs

Table 6.19: Bivariate results for the Independent vs. ê-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	1,060	1,185	0	.69	.69

For VTAs, there were only two instances of bivariance: in the first, NDA.relations.actor always occurred with D.actor. In the Second, third-person actors and goals never occurred together. D.actor and Actor.3 were removed to produce the following set of variables for multivariate analysis: 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.obv, Actor.2, Goal.3, Goal.obv, NA.persons.goal, Px3sg.goal, and Px3pl.goal.

6.2.3 Conjunct-Type

Intransitive Inanimate Verbs

Table 6.20: Bivariate results for the Conjunct-Type alternation: VIIs.

category1	category2	N1	N2	N12	uc.12	uc.21
Sg.actor	NI.object.actor	134	119	110	.60	.64

The single instance of bivariance for the VII class, as seen in Table 6.20, in the Conjunct-Type alternation was the relationship between Sg.actor and NI.object.actor, where the latter nearly always occurred alongside the former. Removing the Sg.actor produced the variable set for multivariate analysis: PV.time, II.sense, II.natural.land, II.weather, and NI.object.actor.

Intransitive Animate Verbs

Table 6.21: Bivariate results for the Conjunct-Type alternation: VAIs.

category1	category2	N1	N2	N12	uc.12	uc.21
NA.persons.actor	Pron.actor	545	292	292	.42	.65
NA.persons.actor	Dem.actor	545	149	149	.20	.50
Pron.actor	Dem.actor	292	149	149	.42	.69
Pron.actor	Prox.actor	292	91	91	.24	.59
Pron.actor	Med.actor	292	56	56	.15	.52
D.actor	NDA.relations.actor	164	164	164	1.00	1.00
D.actor	Px1Sg.actor	164	123	123	.67	.83
NDA.relations.actor	Px1Sg.actor	164	123	123	.67	.83
Dem.actor	Prox.actor	149	91	91	.53	.77
Dem.actor	Med.actor	149	56	56	.31	.67

The VAI class' bivariance concerned only actor effects, as seen in Table 6.21. The `Pron.actor` and `Dem.actor` effects always co-occurred with `NA.persons.actor`; `Dem.actor`, `Prox.actor`, and `Med.actor` were always accompanied by `Pron.actor`; the `NDA.relations.actor` and `Px1Sg.actor` variables always co-occurred with `D.actor`; and `Px1Sg.actor` was always accompanied by `NDA.relations.actor`. Finally, both `Prox.actor` and `Med.actor` always co-occurred with `Dem.actor`. Removing `Prox.actor`, `Pron.actor`, `Dem.actor`, `Med.actor`, `D.actor`, and `Px1Sg.actor` resulted in the following variables to be used in multivariate analysis: `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`, and `Actor.2`.

Transitive Inanimate Verbs

Table 6.22: Bivariate results for the Conjunct-Type alternation: VTIs.

category1	category2	N1	N2	N12	uc.12	uc.21
NA.persons.actor	Pron.actor	162	88	88	.44	.69
Actor.3	Actor.1	1,195	693	0	.58	.62
Dem.goal	Pron.goal	216	216	216	1.00	1.00
Dem.goal	Prox.goal	216	122	122	.46	.68
Dem.goal	Med.goal	216	94	94	.34	.62
Pron.goal	Prox.goal	216	122	122	.46	.68
Pron.goal	Med.goal	216	94	94	.34	.62

The bivariance for VTIs in the Conjunct-Type alternation concerned mostly goal-related variables, as shown in Table 6.22. The variables `Pron.goal`, `Prox.goal`, and `Med.goal` always co-occurred with `Dem.goal`. Both `Prox.goal` and `Med.goal` co-occurred with `Pron.goal`. Removing `Pron.actor`, `Actor.1`, `Dem.goal`, `Pron.goal` resulted in the following set of effect for multivariate analysis: `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`, and `D.goal`.

Transitive Animate Verbs

Table 6.23: Bivariate results for the Conjunct-Type alternation: VTAs.

category1	category2	N1	N2	N12	uc.12	uc.21
Actor.3	Goal.3	893	897	0	.65	.65

The VTA class had only a single bivariate pair, shown in Table 6.23. In this pair, `Actor.3` and `Goal.3` never occurred together. Thus, these two variables 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 that will be used in multivariate analysis: `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`, and `Prox.goal`.

6.3 Multivariate Results

The following section details the results of the multivariate logistic regressions described in Chapter 5. 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). Note that, where relevant, semantic class was used as a fixed-effect. This was done in order to determine the extent to which each semantic-class effect affects the choice of Order, as opposed to the general effect that the variation between semantic classes does. In all cases, results include 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[3] 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 modelling. As Agresti (2013, 165) points out, the intercept is not usually of much explanatory value in and of itself, though to calculate probability estimates, it is necessary. Each effect is reported with an estimate of impact (in log-odds) of the associated effect, as well as a *p* value. A summary table is given for each of the four verb classes in each of the three alternations being studied.

For multivariate results, colour-coding is used to more easily identify which outcome a variable associated with. This was not done previously as Univariate Results included SPR signs ($-$ and $+$) indicating the same information. Because the magnitude of effects (indicated by the reported coefficients) is of importance, similar signs cannot be used in the same way. For binary alternations, a green cell indicates an effect (that is, the `TRUE` value of a variable) increasing the likelihood of an Independent form, while a red cell represent the opposite. For the multinomial alternation, a green cell indicates that an effect increased the likelihood of the outcome indicated by the column name; a red cell represents a decrease for the same.

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

Note that, while the univariate and bivariate results indicated that `PV.time` was a variable worth including in multivariate analyses, it was not used in any inflectional class for the Conjunct-Type alternation. This is because the ka-Conjunct necessarily contains the preverb {ka-}, which is the same morph used for the future definite forms in the Independent. The two uses of this morpheme were not differentiated in the corpus when analyses were run. As a result, any attempt to include the `PV.time` variable may affect results for the Conjunct-Type alternation, because every instance of a ka-Conjunct form (as part of the Other-Conjunct class) would be associated with the `TRUE` value of the variable. To address this, and because {ka-} is not normally a valid preverb of time in the Conjunct, the `PV.time` variable was simply not included in this analysis.

6.3.1 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 was positive, a verb was more likely to occur in the Independent when observed with the `TRUE` value of the variable and less likely to do so for the Conjunct. On the other hand, if an effect was negative, a verb was less likely to occur in the Independent when observed with the variable's `TRUE` value, and more likely to do so in the Conjunct.

Intransitive Inanimate Verbs

Table 6.24: Multivariate results for the Independent vs. Conjunct alternation: VIIs.

	Independent	
	Estimate	*p* value
(Intercept)	-1.344	$< .001$
`PV.time`	0.686	.001

In the alternation between the Independent and the Conjunct, the VII model generally had only a single significant effect: preverbs of time, which increase the likelihood of an Independent form.[4] In fact, of the 204 Independent forms, 81 contained a preverb of time, the vast majority of which were the past tense `PV.ki`, as in (15) and (16). This effect is described in Table 6.24.

The citation in (15) of *C2GB 32*, and similar citations in interlinear glosses throughout this chapter, refers to the corpus file and line number(s) in said file which contains the featured example. See the prefatory *Corpus Abbreviations* page in the front matter of this dissertation to determine which corpus code is associated with which corpus file and the relevant publication that corresponds to each corpus file. Where official translations[5] are available, these examples are given along with a citation identifying a published volume where they can be found.[6]

(15) "êy, **kî-miywâsin**," itwê-w, nôcikwêsiw ana ... (C2GB 32)
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., 1992/1998, 74–75)

(16) **kî-âyiman** ôtê ka-pê-wîcihiwê-yân maskwacîsihk ... (EM 8)
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 conforms to the description of the Order in Cook (2014).

[4]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.

[5]That is, those translations that are given by the editor of the volume associated with the files in which the example was found.

[6]For examples where an official translation could not be found, I have opted not to provide a gloss, due to my lack of fluency in the language. Because these examples are to illustrate the morphosyntactic effects, examples are still presented as they contain such information.

Intransitive Animate Verbs

Table 6.25: Multivariate results for the Independent vs. Conjunct alternation: VAIs.

	Independent	
	Estimate	*p* value
(Intercept)	−1.379	< .001
`PV.discourse`	−0.944	.001
`Actor.obv`	−0.812	.001
`Sg.actor`	−0.472	.003
`PV.time`	0.182	.01

For the Independent vs. Conjunct VAI model, 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 was 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. These effects are given in Table 6.25.

(17) êkwa, wîhkât nânitaw **kâ-isi-mâyinikêhkâto-cik** ôki nêhiyawak ... (VDC2 180)
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 & Douquette, 1987, 46–47)

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 2,157 tokens, to be exact (note that this ratio was much higher than in the Conjunct, where it is 209 of 4,180 tokens). Despite this, the verb class `AI.Speech` was not found to be a significant effect on the Independent Order.

Transitive Inanimate Verbs

Table 6.26: Multivariate results for the Independent vs. Conjunct alternation: VTIs.

	Independent	
	Estimate	*p* value
PV.discourse	−2.064	$< .001$
TI.money.count	−1.914	.02
NI.place.goal	−1.441	.03
NDI.body.goal	−1.059	.04
Actor.3	−0.793	$< .001$
TI.do	−0.766	$< .001$
NI.nominal.goal	−0.752	.010
Actor.2	0.372	.02
NA.persons.actor	0.495	.001
Px1sg.goal	1.613	.005

The Independent vs. Conjunct VTI model had many significant effects, as seen in Table 6.26. As with the other classes, the majority of these effects showed influence towards a Conjunct form. Again the Conjunct Order was associated 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 was more likely to have an overt goal and a non-local actor, as in (19).

(18) ... **ki-kiskêyihtê-nâwâw** kîstawâw ... (AA 20)
2.PL.know.it-2.PL 2.PL.also

'... **you all know this** ...' (Ahenakew, 2000, 40–41)

(19) ... kayâs ayis ês wiyat-~ **ê-kî-wêpina-hkik** wiyat-~
long ago for evidently wiyat-~ CNJ-PST-throw.away-3.PL wiyat-~

wîwatiwâwa ... (AA 89)
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, overt goals of any kind were not significant for the Independent Order. Perhaps the most striking aspect of these results, however, 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 6.27: Multivariate results for the Independent vs. Conjunct alternation: VTAs.

	Independent	
	Estimate	*p* value
`TA.food`	−1.723	.007
`PV.position`	−1.026	.01
`Actor.obv`	−0.921	< .001
`PV.move`	−0.612	.005
`Sg.actor`	−0.608	.04
`Goal.2`	−0.487	.03
`NA.persons.goal`	−0.352	.01
`PV.time`	−0.342	< .001
`Goal.obv`	−0.314	.03
`Actor.1`	0.485	< .001

For the Independent vs. Conjunct VTA model, verbs which regarded food strongly motivated Conjunct forms. Preverbs of position, 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 VTA model: that of first-person actors. In this class, it seems the Conjunct Order is non-present in nature, as well as being modified by preverbs. The Independent still associated with a local actor, but not second-person. These effects are shown in Table 6.27. That the Conjunct was associated with the obviative fits with the VII and VAI classes.

In (20) 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.

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

'... she killed rabbits and **tried to feed us**...' (Bear et al., 1992/1998 208–209)

As noted, the *only* significant effect for the Independent in the Independent vs. Conjunct VTA model was first-person actors. This discrepancy might be written off as an issue of data sparsity, as there were 1,071 Independent TAs in this alternation and 1,931 Conjunct forms, though this difference in data size is not large enough that one would expect numerous effects to be significant for the Conjunct, while only one was for the Independent.

6.3.2 Independent vs. ê-Conjunct

As in the previous alternation, a positive effect influences the production of an Independent form. Negative effects in this alternation represent an increase in likelihood of the ê-Conjunct form specifically.

Intransitive Inanimate Verbs

Table 6.28: Multivariate results for the Independent vs. ê-Conjunct alternation: VIIs.

	Independent	
	Estimate	*p* value
(Intercept)	−0.932	< .001
`PV.time`	0.654	.003

As in the alternation between the Independent and the general Conjunct, the Independent vs. ê-Conjunct VII model had only one significant effect, shown in Table 6.28. The single effect was that of preverbs of time, which increased the likelihood of observing an Independent form. An example of an utterance with such a verb is seen in (21)

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

OFFICIAL TRANSLATION UNAVAILABLE

Intransitive Animate Verbs

Table 6.29: Multivariate results for the Independent vs. ê-Conjunct alternation: VAIs.

	Independent	
	Estimate	*p* value
(Intercept)	−1.529	< .001
`PV.discourse`	−1.087	< .001
`PV.time`	0.204	.008
`Actor.3`	0.373	.007
`Actor.1`	0.632	< .001

In the Independent vs. ê-Conjunct VAI model, discourse preverbs again strongly increased the likelihood of ê-Conjunct forms, while all other significant effects, preverbs of time and first or third-person actors, increased the likelihood of the Independent Order. This is presented in Table 6.29. These results suggest again an Independent form which is more focused on simple declarative structures that may be in a tense other than the present, as in (22), where the main verb *kî-atoskêw* takes an Independent form.

(22) 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 EM 160
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, 102–103)

There are no significant effects in the form of semantic classes of verbs. This was true even in spite of the fact that nearly 21% of all Independent VAIs were forms of *itwêw*, 's/he says.' This quotative was tagged as an `AI-Speech` verb, yet this effect was not found to be significant.

Transitive Inanimate Verbs

Table 6.30: Multivariate results for the Independent vs. ê-Conjunct alternation: VTIs.

	Independent	
	Estimate	*p* value
`PV.discourse`	−2.366	< .001
`TI.money.count`	−2.029	.02
`NI.place.goal`	−1.604	.02
`TI.do`	−.912	< .001
`Actor.3`	−.797	< .001
`NA.persons.actor`	0.583	.001
`Actor.2`	0.851	< .001
`Px1sg.goal`	1.275	.04

For the VTI model in the Independent vs. ê-Conjunct alternation, whose effects are seen in Table 6.30, preverbs of discourse again strongly increased the likelihood of an ê-Conjunct form. Unlike the previous verb classes, the VTI model 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.

(23)

êkoni	kahkiyaw	ê-kî-wâpahta-mân	tânis	âya	**ê-kî-isi-pamina-hkik**	kîkway	... (EM 146)
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, 96–97)

In (23), *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* was 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 this token of *wâpahtam* to occur in the Independent.

Transitive Animate Verbs

Table 6.31: Multivariate results for the Independent vs. ê-Conjunct alternation: VTAs.

	Independent	
	Estimate	*p* value
TA.food	−1.648	.003
PV.position	−1.342	.001
PV.discourse	−1.108	.003
Actor.obv	−0.801	.005
PV.move	−0.506	.03
PV.time	−0.331	.001
Actor.1	0.539	.001
Actor.2	1.807	$< .001$

The Transitive Animate Verb class' significant effects are given in Table 6.31. A semantic class relating to food and preverbs of discourse strongly increased the likelihood that a verb would 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 strongest effect, increasing the likelihood of the Independent. This again suggests the ê-Conjunct as a form associated with discursively marked and less-proximate actions, as well as those displaced in time. This is reflected in (24), where the main verb, *ê-kî-ayi-mâkohikot* is given in the ê-Conjunct.

(24) iyikohk mâna **ê-kî-ayi-mâkoh-iko-t** anihi wîhtikowa tâpwê ... (AA 10)
so much truly CNJ-PST-ah-is.pressed.upon-INV-3SG.3OBV that windigo.OBV truly

'And **he [nêwâpisk] was truly pressed upon** by that windigo ...' (Ahenakew, 2000, 34–35)

6.3.3 Conjunct-Type

The final alternation detailed in this section is multinomial: Conjunct-Type forms. As a result, while a positive effect for an ê-Conjunct outcome represents an increased likelihood of ê-Conjunct forms, a negative value can not be interpreted as an increase in likelihood 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 an OVR 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 parentheses. Green cells indicate that an effect increases the likelihood of an outcome, while red cells indicate that an effect decreases the likelihood of an outcome.

Intransitive Inanimate Verbs

Table 6.32: 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 (< .001)	−2.452 (< .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, as shown in Table 6.32.

(25) mâka mân ânohc **kâ-kîsikâ-k** kâ-mâmitonêyihtamân ... (C2GB 29)
but used to today CNJ-today-3.SG I think about it

‘But when I think of it **today** ...’ (Bear et al., 1992/1998, 218–219)

(26) ... âta **kâ-kimiwa-hk** ... (EM 65)
although CNJ-rains-3.SG

‘... even when **it was raining** ...’ (Minde, 1997, 36–37)

In (25), *kâ-kîsikâk* is used adverbally as an adjunct of time. The kâ-Conjunct here appears to represent a non-hypothetical conditional form, as opposed to the relativized form as in other instances. In (26), the kâ-Conjunct is used conditionally in the past, 'when it was was raining.'

Intransitive Animate Verbs

Table 6.33: 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 (< .001)	−1.342 (< .001)	−3.052 (< .001)
`Actor.2`	−1.147 (< .001)		1.849 (< .001)
`Sg.actor`	−0.695 (.001)	0.633 (.004)	
`Actor.1`	0.471 (< .001)	−0.553 (< .001)	
`NDA.relations.actor`	0.561 (.03)		−2.744 (.01)
`Actor.3`	0.563 (< .001)	−0.653 (< .001)	
`Rdplw`	0.616 (.03)		
`PV.discourse`	0.791 (.003)	−0.719 (.02)	
`PV.position`	1.101 (.001)	−0.985 (.01)	
`PV.wantcan`		−1.227 (.05)	1.830 (< .001)
`PV.qual`			−1.637 (.03)
`AI.health`			1.359 (.005)
`NA.persons.actor`		0.379 (.04)	
`Pl.actor`		0.578 (.02)	

As previously, effects were far more numerous for the Conjunct-Type VAI model than the VII model, as reported in Table 6.33. Second-person actors decreased the likelihood of the ê-Conjunct, but increased 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 last 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 positively affects the ê-Conjunct is 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.

(27) ... êkosi namôya ki-kiskêyihtê-nânaw tânitê **ê-isi-pimohtê-cik**
so NEG PST-know-21.PL where CNJ-thus-walk-3.PL

êkwa kitôskâyiminawak ... (VDC2 151–153)
and our kids

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

(28) â, anohc kâ-kîsikâ-k, êwak ôhc êtikwê ayisiyiniw **kâ-maskawâtisi-t** ... (C2GB 29)
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., 1992/1998, 364)

(29) tânisi k-êtôtamân, **mêstohtê-yêko** pê-miyi-kawi-yâni wêpinâson ... (JK 160)
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 ...' (Kâ-Nîpitêhtêw, 1998, 132–133)

In this alternation as in others, the ê-Conjunct was associated with first and second-persons and those with preverbs of discourse and position, as seen in (27). 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 (28). The effects of the Other-Conjunct did 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 (29).

Transitive Inanimate Verbs

Table 6.34: 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 (< .001)	−2.361 (< .001)	−2.652 (< .001)
`Actor.2`	−1.21 (< .001)		1.847 (< .001)
`Prox.goal`	−1.039 (< .001)	0.898 (.001)	
`PV.wantcan`	−1.031 (.003)	1.365 (< .001)	
`TI.speech`	−0.776 (.02)	0.817 (.02)	
`NI.nominal.goal`	−0.753 (.007)		0.860 (.01)
`Sg.goal`	−0.479 (.02)		
`PV.position`	2.362 (.002)	−2.190 (.04)	−2.203 (.05)
`NA.persons.actor`		0.791 (.002)	
`NI.object.goal`			−0.988 (.004)
`Med.goal`			0.983 (.01)

Significant effects for the VTI model in the Conjunct-Type alternation are shown in Table 6.34. 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/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 to 'position' (metaphorically extended to indicated negativity as previously described), as in (30); that the kâ-Conjunct is more likely to have a proximal goal, have a person actor and *not* have a position preverb, as in 31); and the Other-Conjunct outcome has a second-person actor, a medial goal, but *not* a position preverb, as in (32).

(30) ... môy **ê-ohci-kaskihtâ-yâhk** ka-kîsowihkaso-yâhk ... (EM 191)
NEG CNJ-PST.NEG-be.able-1.PL CNJ-get.warm-1.PL

'... and **we did not manage** to get warm ...' (Minde, 1997, 116–117)

(31) ôhi wiya kayâhtê ayisiyiniwak **kâ-kî-âpacihtâ-cik** ... (AL 684)
this that before people CNJ-PST-use-3.PL

'The things that **people used to use** formerly ...' (Bear et al., 1992/1998, 294–295)

(32) ... kwayask êkwa anita **ta-kakwê-pimipayihtâ-yêk** anima kâ-nêhiyawê-yêk ... (SW 15)
right and there CNJ-try-keep.up-2PL this CNJ-speak.cree-2.PL

'... **you should make a serious effort** to keep speaking your Cree ...' (Whitecalf, 1993, 26–27)

Transitive Animate Verbs

Results for the VTA Conjunct-Type model were similar to those for the VTI model and are given in Table 6.35. 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 but had a strong negative effect for the Other-Conjunct outcome. Goals possessed by singular first-persons

Table 6.35: 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 (.02)
`Actor.2`	−1.595 (< .001)		1.829 (< .001)
`Prox.actor`	−0.938 (.02)	1.324 (.001)	
`Sg.goal`	−0.460 (.04)	0.585 (.01)	
`Actor.1`	0.432 (.007)		−0.751 (.001)
`Px1sg.goal`	0.695 (.05)		
`PV.discourse`	1.359 (< .001)		−2.463 (.02)
`PV.position`	1.775 (.004)	−1.459 (.05)	
`TA.cognitive`		−0.428 (.05)	0.667 (.02)

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 but 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, similar to the way the outcome is framed in the Independent vs. ê-Conjunct alternation. This is exemplified in (33). The kâ-Conjunct class had fewer positive effects, but a verb with a proximate singular actor (and lacking a position preverb) is presented in (34). Finally, the Other-Conjunct class as embodied by second-person actors on verbs of cognition and a lack of discourse preverb is presented in (35).

(33) ... êkosi piko **ê-isi-wîhtamâ-t-akok** ... (JK 21)
that is all only CNJ-thus-tell-INV-1SG.2PL

'... that is all I am telling you ...' (Kâ-Nîpitêhtêw, 1998, 66–67)

(34) 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)
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)

(35) ... **ka-kitâpamâ-yêkok** iskwêwak ôtê ê-sâkaskinêkâpawi-cik ... (JK 144)
CNJ-look.at-2PL.3PL women over there CNJ-stand.crowded-3PL

'... for you to watch these women standing crowded over there ...' (Kâ-Nîpitêhtêw, 1998, 126–127)

6.4 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; as well as measures of classification, precision, recall, τ_c (a measure of how much better a model classifies based on the knowledge of some predictor(s) than simply basing classification off of the overall distribution of a predicted outcome. See Subsection 5.5 for more information), and pseudo-R^2 (ρ^2) (a measure of reduction in 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 such as 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).

6.4.1 Independent vs. Conjunct

Table 6.36: Model statistics for the Independent vs. Conjunct alternation: VIIs.

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	562	44
IND	141	46
Accuracy	77%	
τ_c	.35	
ρ^2	.13	
	CNJ	IND
Recall	93%	25%
Precision	80%	51%

Table 6.37: Model statistics for the Independent vs. Conjunct alternation: VAIs.

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	3,973	282
IND	1,045	1,037
Accuracy	79%	
τ_c	.53	
ρ^2	.27	
	CNJ	IND
Recall	93%	50%
Precision	79%	79%

Table 6.38: Model statistics for the Independent vs. Conjunct alternation: VTIs.

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	1,982	176
IND	561	336
Accuracy	76%	
τ_c	.42	
ρ^2	.16	
	CNJ	IND
Recall	92%	38%
Precision	78%	66%

Table 6.39: Model statistics for the Independent vs. Conjunct alternation: VTAs.

	$\widehat{\text{CNJ}}$	$\widehat{\text{IND}}$
CNJ	1,765	207
IND	541	494
Accuracy	75%	
τ_c	.45	
ρ^2	.21	
	CNJ	IND
Recall	90%	48%
Precision	77%	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 .12 similarly suggests a middling model, as did a τ_c measure of .35.

The model showed similar accuracy (79%), Conjunct recall (93%), and a Conjunct precision (79%), though Independent recall was higher at 50% and 79%. The τ_c and ρ^2 measures were also notably higher than in other models: the former at .53 and the latter at .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 model, with an overall accuracy of 76%, a Conjunct recall of 92%, and precision of 78%, with an Independent recall of 38% and precision of 66%. The VTI model's τ_c of .42 suggests an average increase in classification when compared to other models, while the ρ^2 of .16 suggests 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 .71. The τ_c is .45 and the ρ^2 of .21 represents a relatively well-fitting model.

6.4.2 Independent vs. ê-Conjunct

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%. Recall was 92% for the ê-Conjunct and 36% for the Independent. Precision was 74% in the ê-Conjunct and 69% in the Independent. The model's τ_c measure was the lowest of the inflectional classes at .40 and its ρ^2 was similar at .17.

The VAI model showed values higher than seen in the VII in nearly all measures: accuracy was 76%, and recall was 91% for the ê-Conjunct and 54% for the Independent. Precision was 74% for the ê-Conjunct, but was higher for the Independent at 82%. The τ_c measure was .51 (the highest of the four inflectional classes), and the model's ρ^2 was a relatively high .27, representing a large reduction in badness-of-fit.

Table 6.40: Model statistics for the Independent vs. ê-Conjunct alternation: VIIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{IND}}$
ê-CNJ	378	33
IND	131	73
Accuracy	73%	
τ_c	.40	
ρ^2	.17	
	ê-CNJ	IND
Recall	92%	36%
Precision	74%	69%

Table 6.41: Model statistics for the Independent vs. ê-Conjunct alternation: VAIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{IND}}$
ê-CNJ	2,834	266
IND	987	1,170
Accuracy	76%	
τ_c	.51	
ρ^2	.27	
	ê-CNJ	IND
Recall	91%	54%
Precision	74%	82%

Table 6.42: Model statistics for the Independent vs. ê-Conjunct alternation: VTIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{IND}}$
ê-CNJ	1,322	193
IND	468	490
Accuracy	72%	
τ_c	.44	
ρ^2	.20	
	ê-CNJ	IND
Recall	85%	55%
Precision	74%	73%

Table 6.43: Model statistics for the Independent vs. ê-Conjunct alternation: VTAs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{IND}}$
ê-CNJ	1,083	269
IND	409	662
Accuracy	72%	
τ_c	.43	
ρ^2	.22	
	ê-CNJ	IND
Recall	80%	62%
Precision	73%	71%

The VTI model was slightly less effective, with a 72% accuracy. Recall for the ê-Conjunct was 85%, with the Independent being much lower at 55%. Precision was 74% and 73% for the ê-Conjunct and Independent respectively. The model's τ_c score was .44 and the ρ^2 was relatively large at .20, representing a good fit.

The VTA model's accuracy was 72%. Its recall was 80% for the ê-Conjunct and 62% for the Independent. Precision was 73% for the ê-Conjunct and 71% for the Independent. Finally, the VTA model had a τ_c value of .43, similar to the VII and VTI models, but lower than the VAI model's. The VTA model had a ρ^2 of .22, representing a very good reduction in badness-of-fit.

6.4.3 Conjunct-Type

Table 6.44: Model statistics for the Conjunct-Type alternation: VIIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{kâ-CNJ}}$	$\widehat{\text{other-CNJ}}$
ê-CNJ	375	36	0
kâ-CNJ	59	101	1
other	16	5	3
Accuracy	80%		
τ_c	.56		
ρ^2	.34		
	ê-CNJ	kâ-CNJ	Other-CNJ
Recall	91%	63%	13%
Precision	83%	71%	75%

Table 6.45: Model statistics for the Conjunct-Type alternation: VAIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{kâ-CNJ}}$	$\widehat{\text{other-CNJ}}$
ê-CNJ	3,062	31	7
kâ-CNJ	733	76	10
other	254	4	55
Accuracy	76%		
τ_c	.42		
ρ^2	.21		
	ê-CNJ	kâ-CNJ	Other-CNJ
Recall	99%	09%	18%
Precision	76%	69%	76%

Table 6.46: Model statistics for the Conjunct-Type alternation: VTIs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{kâ-CNJ}}$	$\widehat{\text{other-CNJ}}$
ê-CNJ	1,475	24	16
kâ-CNJ	261	46	13
other	218	9	50
Accuracy	74%		
τ_c	.43		
ρ^2	.25		
	ê-CNJ	kâ-CNJ	Other-CNJ
Recall	98%	14%	18%
Precision	76%	58%	63%

Table 6.47: Model statistics for the Conjunct-Type alternation: VTAs.

	$\widehat{\text{ê-CNJ}}$	$\widehat{\text{kâ-CNJ}}$	$\widehat{\text{other-CNJ}}$
ê-CNJ	1,317	24	11
kâ-CNJ	326	63	7
other	151	20	43
Accuracy	73%		
τ_c	.42		
ρ^2	.20		
	ê-CNJ	kâ-CNJ	Other-CNJ
Recall	97%	16%	20%
Precision	73%	59%	71%

The VII model had an accuracy of 80%. Precision scores were 83% in the ê-Conjunct, 71% in the kâ-Conjunct, and 75% in the Other-Conjunct. Recall was substantially smaller than precision, except in the ê-Conjunct. The ê-Conjunct had a recall of 91%, the kâ-Conjunct had a recall of 63%, and the Other-Conjunct had a very low recall of only 13%. Precision varied far less, with precision rates of 83%, 71%, and 75% for the ê-Conjunct, kâ-Conjunct, and Other-Conjunct, respectively. The VII model's τ_c was .56 and it's ρ^2 was .34, suggesting a well-fitting model.

The VAI model had an accuracy of 76%. Precision scores were 76%, 69%, and 76% for the ê-Conjunct, kâ-Conjunct, and Other-Conjunct respectively. Recall was similar to the VII model. The ê-Conjunct had a very high recall of 99%, the kâ-Conjunct had a very low recall of 9%, and the Other-Conjunct had a low recall of only 18%. Precision for the VAI model was close to that of the VII results, though the ê-Conjunct was lower. The ê-Conjunct outcome had a precision rate of 76%, the kâ-Conjunct had a precision rate of 69%, and the Other-Conjunct had a precision rate of 76%. The τ_c value of .42 and ρ^2 value of .21 suggest a well-fitting model, though one less well-fitting than the VII model.

The VTI model had an accuracy of 74%. Precision scores were 76%, 58%, and 63% for each of the ê-Conjunct, kâ-Conjunct, and Other-Conjunct. Once again, recall was very high in the ê-Conjunct at 98% and low in the kâ-Conjunct and Other-Conjunct with rates of 14% and 18% respectively. Precision was generally lower than in the two previously discussed models, with the ê-Conjunct's precision being 76%, the kâ-Conjunct's being 58%, and the Other-Conjunct's being 63%. The VTI model had similar τ_c and ρ^2 values as the VAI model, with the VTI model having a τ_c score of .43 and a ρ^2 of .25.

The VTAs had an accuracy of 73%. Precision scores of 73%, 59%, and 71% for the ê-Conjunct, kâ-Conjunct, and Other-Conjunct respectively. Recall in this model followed the trend of a very high ê-Conjunct rate at 97% followed by a kâ-Conjunct rate of 16%, and an Other-Conjunct rate of 20%. Similar to the VTI model, the VTA model's ê-Conjunct precision was 73%, the kâ-Conjunct's was 59%, and the Other-Conjunct's was 71%. Finally, the VTA model had similar model fit statistics as the VAI and VTI model, with a τ_c value of .42 and a ρ^2 value of .20. This again suggests a well-fitting model.

6.4.4 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_c

scores (as discussed above) to determine the relative importance 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_c 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 include 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 with the random-effect of `Lemma` (abbreviated SEM.LEM), those which included the random-effect of `Lemma` but had only *morphological* effects as fixed-effects (abbreviated MORPH.LEM), and those full mixed-effect models (abbreviated ME) which include both types of fixed-effects and the random-effect.[7] Further, there are two additional models presented, each without random-effects: models with only semantic-effects (abbreviated SEM) and models with only morphological-effects (abbreviated MORPH). Table 6.48 depicts the composition of the models to be compared.

Table 6.48: Model composition.

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

[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.

For this dissertation, morphological-effects are those which have obvious and easily identifiable morphological exponents, such as `Rdplw` 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-effects used in any model:

```
Actor.1
Actor.2
Actor.3
Actor.obv
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.food.actor

NA.persons.actor

NA.persons.goal

NDA.relations.actor

NDI.body.goal

NI.natural.force.goal

NI.nominal.goal

NI.object.actor

NI.object.goal

NI.place.goal

PV.discourse

PV.move

PV.position

PV.qual

PV.startfin

PV.time

PV.wantcan

Prox.actor

Prox.goal

TA.cognitive

TA.do

TA.food

TA.money.count

TA.speech

TI.do

TI.money.count

TI.speech

Table 6.49 presents the τ_c and ρ^2 values in each of the five different types of models previously described for the Independent vs. Conjunct alternation.

Table 6.49: Model comparisons. Independent vs. Conjunct alternation. Green cells with bold items represent a very good model fit, per McFadden (1973).

	VII		VAI		VTI		VTA	
	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2
SEM.MORPH	.31	.05	.47	.15	.32	.09	.35	.10
LEM	.36	.12	.52	**.26**	.34	.15	.42	.18
SEM.LEM	.35	.12	.52	**.27**	.36	.15	.44	.19
MORPH.LEM			.52	**.27**	.39	.16	.43	.19
ME	.35	.12	.53	**.27**	.42	.16	.45	**.21**
SEM	.31	.05	.47	.14	.30	.06	.39	.08
MORPH			.26	.03	.28	.04	.24	.04

In this alternation, ME models often had superior performance in both classification improvement (τ_c) 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_c and ρ^2 than the ME model, despite containing less information in terms of predictors. It is worth noting that the VII infectional class was the least-numerous class, and due to its inherent semantics, it is substantially different than the other classes (in that it can refer to things like days of the week and temporal states). Also worth mentioning is that MORPH.LEM models were not available for the VII class, as the significant fixed-effects for the VII model were all semantic. The other case is in the VTI model where the ρ^2 was the same for ME models as the MORPH.LEM models. In all cases, SEM.MORPH models exhibited much lower measures than other models, with LEM models showed a substantial increase in ρ^2 and τ_c measures over SEM.MORPH models. This indicates that a substantial amount of alternation is explained by random-effects only. Put another way, individual lemmas appeared to show substantial 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 showed a slight increase in ρ^2 and τ_c over the LEM models. While SEM.LEM and MORPH.LEM varied in which produces a better model depending on the verb inflectional class, this difference was usually minimal. Interestingly, all models except SEM.MORPH performed extremely similarly in the VII and VAI classes, indicating that the influence of a random-effect (lemma) is one of, if not the, most important factors in modelling this alternation in these inflectional classes. The SEM and MORPH models without random-effects were always the worst-performing models for both measures, especially when compared with their random-effect counterparts. As mentioned before, the ME models often perform better than any other model (unsurprisingly), and, excluding the VII model, they were never *worse* than the other models in their ability to classify or reduce badness-of-fit.

Table 6.50: Model comparisons. Independent vs. ê-Conjunct alternation. Green cells with bold items represent a very good model fit, per McFadden (1973).

	VII		VAI		VTI		VTA	
	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2
SEM.MORPH	.28	.04	.43	.15	.33	.10	.32	.12
LEM	.36	.14	.49	**.26**	.36	.17	.38	.17
SEM.LEM	.40	.17	.50	**.27**	.39	.17	.41	.19
MORPH.LEM			.50	**.26**	.41	.19	.44	**.21**
ME	.40	.17	.51	**.27**	.44	**.20**	.43	**.22**
SEM	.28	.05	.43	.14	.24	.07	.31	.09
MORPH			.15	.03	.24	.05	.23	.08

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 susbtantially higher ρ^2. This again suggests that lemmas have inherent propensities to surface in one Order over another. Again, models without random-effects were always substantially lower-performing than models with. In all cases, ME models were the best-fitted models, except for the ME VAI model which was was equal to the SEM.LEM VAI model. Similarly, all classes other than the VII showed ME models with a ρ^2 greater than .20. The lower measure for the VII ME model is again likely the result of a paucity of data.

Taken together, the model performance for the Independent vs. ê-Conjunct alternation suggests a generally better-modelled 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 clearer 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 types of Conjunct forms. Cook (2014) describes the ê-Conjunct as more of an elsewhere case and a Conjunct form similar to the Independent in morphosyntactic behaviour. In fact, Cook (2014, 125) describes all types of Conjunct other than the ê-Conjunct as being disallowed from matrix clauses, which are the domain

Table 6.51: Model comparisons. Independent vs. ê-Conjunct alternation with results of Chapter 4.

	VII	VAI	VTI	VTA
Semi-Automatic	.18	.13	.04	.06
HAC-Only	.19	.09	.01	.06
SEM	.05	.14	.07	.09

of the Independent and the ê-Conjunct. This comports with the results of the mixed-effects models. Further, the forms in the Other-Conjunct category 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, considering the ê-Conjunct together with these other, more straightforwardly described forms (as is done in the Independent vs. Conjunct alternation), may produce an outcome that a logistic model is not fully able to reproduce and explain.

The SEM model for the Independent vs. ê-Conjunct alternation may seem analogous to the model presented in Chapter 4 (which was assessed for the same alternation); however, there are important differences. The most notable difference is the fact that the models from Chapter 4 were fitted using semantic classes only as random-effects, while the models in this chapter used semantic classes as fixed-effects. This was done to evaluate the effect of particular semantic classes rather than the use of semantic classes generally, as in Chapter 4. Further, models from Chapter 4 were fitted using only the verb/noun classes that were semi-automatically generated. The SEM model detailed in this chapter makes use of other semantic-effects, such as those detailed above. Consequently, the results are not directly comparable. That said, a juxtaposition of the SEM models and the results of Chapter 4 is given in Table 6.51. The SEM models fitted for this section (in the Independent vs. ê-Conjunct alternation as was modelled in the previous chapter) performed better than either the semi-automatic or HAC-only verb class groups from Chapter 4. The only exception to this is in the VII model, where the results were much lower for the SEM model than either the semi-automatic or fully-automatic classes.

These results suggest that the additional information included in the models from this chapter (i.e., everything other than the semi-automatic verb and noun classes) provided useful and important input over just using verb and noun class, except in the VII class. It is unclear why this might be. Given this and other differences between the two sets, future research should further investigate the differences between the VII and other infectional classes.

Table 6.52: Model comparisons. Conjunct-Type alternation. Green cells with bold items represent a very good model fit, per McFadden (1973).

	VII		VAI		VTI		VTA	
	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2
SEM.MORPH	.33	.11	.37	.05	.38	.07	.36	.07
LEM	.55	**.35**	.37	.18	.40	**.20**	.35	.16
SEM.LEM	.56	**.34**	.38	.19	.41	**.23**	.38	.17
MORPH.LEM			.37	.18	.40	**.20**	.35	.16
ME	.56	**.34**	.42	**.21**	.43	**.24**	.42	**.20**
SEM	.33	.11	.36	.02	.38	.04	.34	.02
MORPH			.36	.04	.37	.04	.38	.17

In the Conjunct-Type alternation, as depicted in Table 6.52, models performed variously. All ME models showed substantial reduction in badness-of-fit and improved classification (with ρ^2 of at least .20 and τ_c values of greater than .40). The VAI and VTA models were only successful in the ME case. Conversely, VII and VTI models were successful in all cases except for SEM.MORPH. These values indicate that one requires semantic and morphosyntactic information along with lemma-specific identities to accurately describe this Conjunct-Type alternation. Specifically, it appears that, for the VII and VTI classes, lemma information is paramount for accurate classification. This is more generally reflected in the relative performance of the SEM.MORPH, SEM, and MORPH models, which often resulted in much less well-fitting models when compared with mixed-effects equivalents (though this was less pronounced in the τ_c scores).

6.5 Cross-Validaiton

In addition to reviewing the model statistics, 10-fold cross-validation was undertaken for the main mixed-effects models presented in this dissertation. Cross-validation can be used to judge how generalizable a model is beyond its training data by testing models with reserved test sets. This ensures that models are not being evaluated only with the data they were trained on. To perform cross-validation, k-fold validation as described in James et al. (2013, 181) was chosen as a method. This allows for validation without the need for more data; other comparable techniques such as bootstrapping also exist.

As with other parts of this dissertation, the code to run this cross-validation is available in the online appendix of this dissertation. This cross-validation entailed randomly splitting the data of each model into 10 separate data subsets, also known as *folds*. Each of the 10 folds is selected and placed aside to be used for testing purposes. Using the same formulae as in the ME models of this dissertation, new models are fit using the remaining nine folds as the training data set and evaluated on the reserved test set. This is then repeated until every fold has been reserved for testing and has also been used for training the models (James et al., 2013, 181). This entire process is then repeated for each of the ME models previously fitted in this dissertation.

Model statistics were then run for these models; in this section, a range and mean for each model's accuracy is reported in Table 6.53. Ranges and means for other statistics like ρ^2 and τ_c measures are also given Appendix B. Where ME results fell within the cross-validation (CV) range, this was considered evidence that the models were reasonably in line with the cross-validation set (thus reasonably generalizable and justifiable). Where an ME measure falls within the CV range, the cell is reported normally. Where an ME measure did not fall within the CV range, *and* it differed by more than .03 (or 3 percentage points for accuracy), then it was reported in a red cell. Red cells represent a measure who's generalizability can be called into question. The results from cross-validation for each fold individually are available at `https://github.com/atticusha/DissertationCode/tree/a7eeed9b1036c7201b3c52248eef3d7f0b477b59/Cross-Validation`.

Table 6.53: Cross-validation of the Independent vs. Conjunct models.

	VII	VAI	VTI	VTA
Independent vs. Conjunct				
ME Accuracy	77%	79%	76%	75%
CV Accuracy Mean	81%	78%	77%	73%
CV Accuracy Median	81%	78%	77%	73%
CV Accuracy Range	81%–82%	77%–78%	72%–78%	72%–74%
Independent vs. ê-Conjunct				
ME Accuracy	73%	76%	72%	72%
CV Accuracy Mean	74%	77 %	73%	71%
CV Accuracy Median	75%	76 %	73%	71%
CV Accuracy Range	67%–77%	76%–77%	70%–74%	68%–72%
Conjunct-Type				
ME Accuracy	80%	76%	74%	73%
CV Accuracy Mean	77%	74%	73%	71%
CV Accuracy Median	77%	74%	73%	71%
CV Accuracy Range	66%–86%	72%–76%	67%–78%	67%–75%

All cases but the VII ME model of the Independent vs. Conjunct alternation were within the CV ranges and no more than three percentage points from the CV medians or means. This suggests that the results of the mixed-effects modelling were reasonably generalizable, and valid. See Appendix B for details regarding other cross-validation using other model statistics.

6.6 Exemplar Extraction

In addition to the results presented above, one can make use of logistic regression models in other ways. Although logistic regression models are often thought of as classifiers, they can also be described as estimators. Given an observed data point (i.e., observed token, in the case of this dissertation a verb-form in one of the alternative Orders), the logistic models look at the predictive context provided (e.g., the fixed-effects) as well as the outcome of the data point (i.e., the value of the response variable). After doing so for the entire data set, the logistic model estimates a probability that each data point will be observed with a particular outcome. One can use these estimations to identify examples of predictive contexts that are most likely to produce each outcome. However, similar contexts will produce similar probability estimates, and so selecting only the highest estimates overall will likely result in the same few contexts being repeated. If one wants to learn about the different types of contexts that are most likely to predict an outcome, this is an undesirable situation. Therefore, one might rather want to learn about the variety of different types of contexts and the outcomes they predict as most likely. (Arppe, 2008, 228–252) presents a solution by grouping data points with similar contexts into distinct clusters and then selecting only those data points and then selecting from each bin a single context or small subset of contexts that provide the highest estimated probability for an outcome. This binning approach allows for a diversity of contexts to be represented while still selecting highly-probable estimates. Importantly, only data points where the outcome was correctly predicted were selected (e.g., if a data point has a probability estimate of 75% for $Outcome_a$ when, in fact, it was observed with $Outcome_b$, it would not have been selected). Thus, the selected high-probability exemplary contexts are actually observed ones in all respects.

For this dissertation, I followed the approach of Arppe (2008). Using the models developed in the previous subsections, I extracted the estimated probabilities from the logistic models. Here, data points are each of the fully-inflected verb tokens observed in the corpus (in one of the various Orders studied in this dissertation). The predictive contexts are the presence or absence of the fixed-effects used in modelling. Only those predictors found to be statistically significant were included as part of this context. Token-variable data frames were constructed. In these, each row represents an observed token and each column represents a predictor variable. Cells were binary and represented whether a token was observed with some set of variables. Along with predictor columns was a sentence index identifier, so that the full-sentence context for each token could be extracted. Based on these data frames, automatic clustering of tokens based on predictive context was undertaken with HAC using the Ward method and a binary metric. This analysis was done using the `fviz_silhouette` function from the factoextra library (Kassambara & Mundt, 2020). To ensure the correct number of clusters was used, I used silhouette analysis to evaluate the clustering. I selected the number of clusters that created an average silhouette as close to 1.00 as possible (the average silhouette for all classes divided by the number of classes). This results in some individual classes with low silhouettes, though it ensures the overall clustering is as well-fitting 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 .99) with the lowest number of clusters was chosen. In most cases, an optimal number of silhouettes ranged from 5 to 50.

The result of this analysis is a list of tokens, the associated cluster, a probability for an outcome, and a sentence index. Only those tokens where the predicted outcome (in terms of its Order) matched the actual outcome were considered. Tokens were ordered from most probable to least probable to occur in a particular Order. For each token, the associated sentence was extracted based on the sentence index to produce an *exemplar sentence*. Only the five topmost tokens in terms of the estimated probability for the Order of the verb that actually occurred were chosen as exemplars from that particular bin (and the similar contexts it represents). Table 6.54 details the optimal number of clusters for each alternation and each class.

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

Table 6.54: Number of silhouettes used for clustering.

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

alternative-outcome probabilities in the binary cases will be given in the form of 1−probability, so that an estimate of .01 in the Independent vs. Conjunct alternation (indicating a likely-Conjunct form) will be presented as a Conjunct probability of .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`) of the file along with the line number where the exemplar occurs. A list of corpus codes and the full name of the corpus file, alongside the relevant publication for each 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 for the particular outcome being discussed. In this subsection, 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. Finally, where a tilde (~) is given, it represents a moment of hesitation and is present in the underlying corpus.

6.6.1 Independent vs. Conjunct.

Inanimate Intransitive Verbs

Independent

(36) otâkosihk ma cî wiya **kî-pêhtâkwan** kwayask. (CMBK-5-2 20; .69)
yesterday NEG Q for PST-is.heard properly

OFFICIAL TRANSLATION UNAVAILABLE

(37) aya mîna ni-kî-âtotên kayâs, namôy wîhkâc **ohci-pêhtâkwan** ... (AL 976; .54)
ah and I told about it long ago never **it was heard**

'I have also said it was unheard of long ago ...' (Bear et al., 1992/1998, 318–319)

There were only two cases containing VIIs with probability estimates of over .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 (36), the verb is used in an interrogative clause. Estimated probabilities were 0.54 and 0.69.

Conjunct

(38) "tâpwê anim âkosi sâsay **ê-ispayi-k** anima kâ-kî-itwêt," itwêw ... (C2GB 40; .99)
truly that thus already CNJ-fares.thus-3.SG that s/he said say-3.SG

' "It is true, and some of what he had said **is happening** already," she said ...' (Bear et al., 1992/1998, 80–81)

(39) êkosi anima mîna êwako **ê-kî-ispayi-k** mâna ... (SW 41; .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 .5. Both instances are forms of the lemma *ispayin* 'it happens thus.' In both cases, the predicted verbs seem to carry the main semantics of the utterance, and both appear to be in past tense forms, though in (38) this does not appear to be marked morphologically. This comports with the finding that the Conjunct Order is somehow less immediate, as discussed previously. The Conjunct forms above were well-predicted, with the lowest probability being .97.

Animate Intransitive Verbs

Independent

(40) vamps aniki, âha, 'asêsinwa' **kî-itwê-wak**. (AL 1144; .84)
vamps those yes 'asêsinwa' PST-say-3.PL

'The vamps, yes, **they used to call** them "asêsinwa.' (Bear et al., 1992/1998, 330–331)

(41) ... nititwân mâna, tâspwâw mâna wiya niya **nit-itwâ-n** ... (SW 140; .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, 76–77)

(42) êkwa êkosi **kî-itwê-w** ana kisêyiniw. (VDC2 1061–1062; .79)
and this PST-say-3.SG that old man

'And this is what that old man **said**.' (Vandall & Douquette, 1987, 106–107)

(43) "a play ôm ê-wî-ayâyâhk ôtê Sandy Lake," **itwê-w** ... (AA 33; .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, 44–45)

(44) **ê-kî-wîhkitisit** mâna, ban-~ bannock ê-kî-osîhât ... (C6IC 12; .50)
CNJ-PST-taste.good-3.SG used to ban-~ bannock s/he made it

'The bannock **used to taste good**, and she used to make it ...' (Bear et al., 1992/1998, 148–149)

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 .75 to .84, suggesting decent confidence in the prediction of an Independent outcome. The one lemma that did not concern speech was in (44), where *wîhkitisiw* had a probability of .50, essentially a completely non-confident decision.

Conjunct

(45) ... êkot[a] êkwa ki-kâh-kî-wîcêwâw tânis **ê-isi-mawimoscikê-t**. (JK 8; .97)
there and you would join him/her how CNJ-thus-pray-3SG

'... then you would be able to **join him in his way of worship**.' (Kâ-Nîpitêhtêw, 1998, 50–51)

(46) ... otôsk-âyima êkâ kwayask **ê-isi-wîcêhto-yit**. (JK 6; .97)
young people NEG properly CNJ-thus-join.together-3.OBV

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

(47) ... tânisi **ê-kî-isi-mawimoscikê-cik** nêhiyawak kayâs ... (VDC2 1050–1051; .97)
how CNJ-PST-thus-pray-3PL Cree long ago

'... how the Crees **used to worship** long ago ...' (Vandall & Douquette, 1987, 106–107)

(48) ... 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; .96)
much CNJ-come-call-3.OBV then

'... and then people would realize why **he had come to hoot** there ...' (Whitecalf, 1993, 38–39)

(49) môy mâka wîhkâc ohci-wîhtam-wak awiyiwa anihi
NEG still ever they did not tell about it someone that

kâ-kî-itahkamikisi-yit, êha. (CMBK-5-2 110; .95)
CNJ-PST-behave.thus-3.OBV yes

OFFICIAL TRANSLATION UNAVAILABLE

Unlike the Independent outcome, there appeared to be more variety in regards to what lemmas occurred in the top-five Conjunct exemplar forms. Further, estimated probabilities were much higher, with the model predicting all five exemplars at above a .95 estimated probability. In all cases, these exemplars clearly showed subordination as described by Cook (2014) and Wolfart (1973). For example, (45) contains the exemplar verb, *ê-isi-mawimoscikêt*. Here, *ê-isi-mawimoscikêt* references the way someone prays and is subordinate to the main verb of this clause, *kikâh-kî-wîcêwâw* 'you would be able to join him.' Similarly, for the exemplar seen in (48) 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

(50) "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; .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., 1992/1998, 190–191).

(51) **ki-kiskêyiht-ê-n** kiya? (AL 558; .80)
2-know-THM-2.SG you

'**How about you**?' (Bear et al., 1992/1998, 284–285)

(52) "kiya **ki-kaskiht-â-n** ê-osîhtâ-yan," nit-it-ik. (AA 76; .74)
you 2-be.able-THM-2.SG you make it s/he says to me

' "You, **you have been able** to make it," she says to me.' (Ahenakew, 2000, 68–69)

(53) nîsta **ni-wî-nipahi-cîhkêyiht-ê-n** (CMBK-3-2 182; .76)
I too 1-FUT.VOL-very-like-THM-1.SG

OFFICIAL TRANSLATION UNAVAILABLE

(54) "kiyâm tâpwêhta, môy **ki-ka-mihtât-ê-n**," ê-kî-isit mâna. (CMBK-4-2 114; .74)
please truly NEG 2-FUT.CON-regret-THM-2.SG s/he says to me used to

OFFICIAL TRANSLATION UNAVAILABLE

Independent VTI exemplars mostly had to do with cognition/emotion verbs with the {-êyi-} morph. The exception to this is seen in (52), where a form of *kaskihtâw*, 's/he is able (to do it)' occurs in the Independent. The estimated probabilities were lower than above, ranging from .74 to .80.

Conjunct

(55) 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; 1.00)
s/he makes it pounded meat s/he eats him/her

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

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

ê-isi-osîht-â-t. (C7MW 86; .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., 1992/1998, 192–193)

(57) 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-tipaha-mâhk, mitâtahtomitanaw-maskimot ê-kî-ayâyâhk ... (EM 117; .99)
CNJ-PST-thus-measure-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)

(58) “kây, êkâya mâto! ê-nôhtêhkatêt ana wîst ôm ê-~ ê-wâpamiko-~ ê-wâpahtahk
no NEG 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; .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., 1992/1998, 112–115)

(59) ... ê-wî-nanâskomot, matotisân ôma **kâ-wî-osîht-â-t**. (JK 42; .98)
... s/he gives him/her 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, 1998, 82–83)

The VTI Conjunct exemplars were composed mostly of forms of *osîhtâw*, ‘s/he makes it,’ as seen in examples (55), (56), and (59). These verbs occur in both main and subordinate clauses. Example (57) *ê-kî-isi-tipahamâhk* contains a discourse preverb, while in (58) the exemplar verb, *ê-nôhtê-mîcit*, is desiderative. The Conjuncts were very well-predicted with the lowest estimated probability being .98.

Transitive Animate Verbs

Independent

(60) 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-4-2 304; .79)
we worked

OFFICIAL TRANSLATION UNAVAILABLE

(61) ... ômisi mâna **ni-kî-it-â-wak** nitawâsimisak ... (EM 66; .73)
thus used to 1-PST-say-DIR-1SG.3PL my children

‘... **I used to tell** my children as follows ...’ (Minde, 1997, 36–37)

(62) “îwahikanak niwî-osîhâwak,” **nit-it-â-wak** awâsisak. (AL 407; .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., 1992/1998, 206–207)

(63) “nôsisim!” **nit-it-ik** ... (C2GB 14; .70)
my grandchildren 1-say-INV.1SG.3SG

‘ “Grandchild!” **she said to me** ...’ (Bear et al., 1992/1998, 68–69)

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

'... and the old lady [*sc.* Norman's mother] said to her son, “Let's rest;” ...' (Bear et al., 1992/1998, 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. In all but one instance, seen in (61), these quotatives 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. This was the verb *kinisitohtâtinâwâw*, 'I understand you all.' In this instance, the estimated probability was only .57, representing a relatively uncertain prediction. Generally, estimated probabilities were lower for this group than previously seen, ranging from .63 to .79.

Conjunct

(65) 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; .98)
their young over there

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

(66) ... “kita-wâpamikot,” “**ê-pê-minihkwât-â-yit**,” itwêw. (VDC2 485–486; .97)
looking at him/her CNJ-come-drink-DIR-3.OBV s/he said

'... “looking at him,” he said, “**to trade it for a drink**” he said.' (Vandall & Douquette, 1987, 68–69)

(67) ... wâposwa ê-kî-nipahât **ê-wî-kakwê-asam-iko-yâhk** wiya ... (C8GB 13; .97)
rabbits s/he kills him/her CNJ-FUT.VOL-try-feed-INV-3SG.1PL for

'... she killed rabbits and **tried to feed us...'** (Bear et al., 1992/1998, 208–209)

(68) ... âta tâpiskôc êkâya kîkway wiyasiwêwin wîyawâw
though like nothing law they

ê-ohci-tâwiskâ-ko-cik, nânitaw itinikêtwâwi. (VDC2-RES 233–234; .97)
CNJ-NEG.PST-encounter-INV-3OBV.3FURObv 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 & Douquette, 1987, 50–51)

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

nititikwak. (VDC2-RES 561–562; .97)
they say to me

‘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 & Douquette, 1987, 74–75)

The Conjunct exemplars for the Independent vs. Conjunct VTA set were largely subordinate verbs, such as *ê-pê-minihkwâtâyit*, ‘comes to trade a drink for it’, seen in (66) or *ê-wî-kakwê-asamikoyâhk*, ‘to try to feed all of us’ seen in (67). These results comport with the general descriptions of Order in the literature. Exemplars were all highly predicted, with the estimated probabilities for Conjunct exemplars never being lower than .97.

Note that for (66), the corpus and published source gave quotations in italics instead of using quotation marks. In this section I have used quotation marks to maintain consistency with other exemplars. Further, the corpus differs from the published version in this example, where *itwêw* is given between *kita-wâpamikot* and *ê-pê-minihkwâtâyit* in the published, but not in the raw text file I have used. This example is also used in (102) .

6.6.2 Independent vs. ê-Conjunct

Intransitive Inanimate Verbs

Independent

(70) ... otâkosihk ma cî wiya **kî-pêhtâkwan** kwayask. (CMBK-5-2 20; .85)
yesterday not Q for PST-be.heard.3SG properly

OFFICIAL TRANSLATION UNAVAILABLE

(71) aya mîna ni-kî-âtotên kayâs, namôy wîhkâc **ohci-pêhtâkwan** ... (AL 976; .74)
ah and I told about it long ago never NEG.PST-be.heard.3SG

‘I have also said **it was unheard of** long ago ...’ (Bear et al., 1992/1998, 318–319)

As in the previous alternation, there were fewer than five exemplars available for each outcome in the VII class. There were 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 estimated probabilities were .74 and .85.

ê-Conjunct

(72) "tâpwê anim âkosi sâsay **ê-ispayik** anima kâ-kî-itwêt ..." (C2GB 40; .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., 1992/1998, 80–81)

(73) êkosi anima mîna êwako **ê-kî-ispayik** mâna ... (SW 41; .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)

(74) ê-pânisamihk anima kahkiyaw, nama kîkway **ê-ohci-wêpinikâtê-k**. (CMBK-4-2 250; .50)
someone cuts it that all nothing CNJ-NEG.PST-be.discarded-3.SG

OFFICIAL TRANSLATION UNAVAILABLE

The ê-Conjunct outcome had three exemplars, two of which were forms of the verb *ispayin*, 'it happens.' Seen in (73) and (74) where both verbs are in the past tense. Unlike the Independent outcome, the Conjunct had a large range in estimated probabilities, ranging from .50 to .99.

Animate Intransitive Verbs

Independent

(75) "... môy kîhtwâm êkwa nika-pakitinâw wîhkâc awâsis," **nikî-itwâ-n** ôma ... (CMBK-3-2 170; .90)
NEG again and I will let go ever children 1-PST-say-1SG this

OFFICIAL TRANSLATION UNAVAILABLE

(76) ... nititwân mâna, tâspwâw mâna wiya niya **nit-itwâ-n** ... (SW 140; .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, 76–77)

(77) êkwa êkosi **kî-itwê-w** ana kisêyiniw ... (VDC2 493–494; .88)
and so PST-say-3SG that old man

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

(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; .82)
2-PST-say-2SG

OFFICIAL TRANSLATION UNAVAILABLE

(79) "a play ôm ê-wî-ayâyâhk ôtê Sandy Lake," **itwê-w** ... (AA 33; .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, 44–45)

As with the previous alternation, the VAI Independent exemplars were all forms of *itwêw*, 's/he said.' In all examples other than (76), the exemplar verbs are semantically past tense, 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 .75 to .90.

ê-Conjunct

(80) ... otôsk-âyima êkâ kwayask **ê-isi-wîcêhto-yit**. (JK 7; .97)
their young people NEG right CNJ-thus-cooperate-3.OBV

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

(81) ... êkot[a] êkwa kikâh-kî-wîcêwâw tânis **ê-isi-mawimoscikê-t**. (JK-C4ARC.798 8; .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, 50–51)

(82) êkos êtikwê piko **ê-kî-isi-ma-mêyiwiciskê-hk** ê-kî-isi-pasikôhk. (C8GB 18; .96)
so apparently only CNJ-PST-thus-RDPLW-be.dirty-UNSPEC someone gets up

'... one simply **got up dirty**, I guess.' (Bear et al., 1992/1998, 210–211)

(83) ê-kakwêcimak ôma, tânis **ê-kî-pê-ay-isi-pimâcihiso-cik** ayisiyiniwak ... (AL 2; .95)
I am asking him/her this how CNJ-PST-come-RDPLW-thus-live-3.PL people

'I am asking her this: **how people lived** ...' (Bear et al., 1992/1998, 240–241)

(84) tânitê kiy ê-kî-kiskinahamâkawiyan cî, ka-isi-kakâyawisî-yan **ê-awâsisîwiyan**? (AL 359; .95)
where you you went to school Q you work hard CNJ-be.child-2.SG

‘Where were you taught to work so hard; **when you were a child**?’ (Bear et al., 1992/1998, 270–271)

Nearly all of the exemplars for the Conjunct VAIs made use of the {isi-} preverb, a preverb classified as a discourse preverb, indicating the ways in which an action is done. The only exemplar without this preverb was in (84), the simple second-person ê-Conjunct form, *ê-awâsisîwiyan*, ‘you were a child.’ Estimated probabilities ranged from .95 to .97.

Transitive Inanimate Verbs

Independent

(85) ... mâka ôma **ki-wâpaht-ê-n** ôma niskîsik ... (VDC2 941; .90)
but this 2-see-THM-2.SG this my eye

‘... but **you see** this eye ...’ (Vandall & Douquette, 1987, 98–99)

(86) ... **ki-kiskêyiht-ê-nâwâw** kîstawâw ... (AA 20; .89)
2-know-THM-2.PL you all

‘... **you all know this** ...’ (Ahenakew, 2000, 40–41)

(87) “kiyâm tâpwêhta, môy **ki-ka-mihtât-ê-n** ...” (CMBK-4-2 114; .87)
so agree NEG 2-FUT.CON-regret-THM-2.SG

OFFICIAL TRANSLATION UNAVAILABLE

(88) ... niya wiya namôy **nôh-cîhkêyihtên**. (CMBK-4-2 209; .87)
I for NEG 1.NEG.PST-like-THM-1.SG

OFFICIAL TRANSLATION UNAVAILABLE

(89) ... kîspin ê-sôhkêpayik êkotowahk maskihkiy, namôya, namôya ahpô kikaskihtân
if it is strong like that medicine NEG NEG or you are able

ta-nipâ-yan ... (SW 112; .67)
to CNJ-sleep-2.SG

‘... if that kind of medicine is very strong, you will not even be able **to sleep** ...’ (Whitecalf, 1993, 68–69)

The VTI exemplars in the Independent case were varied, with no lemma being repeated. Two of the exemplar verbs *kikaskihtân* and *kiwâpahtên* were verbs of doing, while all other verbs were those of cognition. Estimated probabilites for this inflectional class ranged from .80 to .90.

(90) 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; 1.00)
s/he makes it pounded meat s/he eats him/her.

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

(91) 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-tipaha-mâhk, mitâtahtomitanaw-maskimot ê-kî-ayâyâhk ... (EM 117; .99)
CNJ-PST-thus-measure-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)

(92) ... mêtoni mân **ê-kî-kanâcihtâ-cik** êkwa mân ê-kî-kaskâpasahkik. (EM 268; .98)
very used to CNJ-PST-clean-3.PL and used to they smoked it

'... they certainly **used to clean** those guts out thoroughly and smoke them.' (Minde, 1997, 140–141)

(93) ... âhci piko pêyakwan iyikohk **ê-kî-isi-môcikêyihta-mihk**. (CMBK-3-2 271; .98)
still only similar until CNJ-PST-thus-UNSPEC

OFFICIAL TRANSLATION UNAVAILABLE

(94) ê-nôhtêhkatêt ana wîst ôm ê-~ ê-wâpamiko-~ ê-wâpahtahk ôma wiyâs
s/he is hungry this one s/he too and ê-~ ê-wâpamiko-~ s/he shows it that meat

ê-nôhtê-mîci-t ... (C4MF 68; .97)
CNJ-want-eat-3.SG

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

All of the Conjunct VTI exemplars for this alternation were past tense, except for (94); further, the majority of these top exemplars, seen in examples (90), (91), and (93) contained the discourse preverb {isi-}. Estimated probabilities were quite high and ranged from .97 to 1.00.

Transitive Animate Verbs

Independent

(95) ... môy êkw êkonik mîna **ki-kî-wîh-â-wak** ... (AL 1284; .91)
NEG and those and 2-PST-rely.on-DIR-2SG.3PL

'... now you can't even **rely on them** ...' (Bear et al., 1992/1998, 342–343)

(96) â, **kit-ayâw-â-wak** cî (AL 106; .86)
ah 2-have-DIR-2SG.3SG Q

'Ah, do **you have any of that**?' (Bear et al., 1992/1998, 250–251)

(97) ... sapiko mân êkosi **nit-it-â-wak** nôsisimak ... (CMBK-4-2 304; .85)
actually used to so 1-say-DIR-1SG.3PL my grandchildren

OFFICIAL TRANSLATION UNAVAILABLE

(98) ... **itâ-wak** mân ôki niwâhkômâkanak ... (EM 160; .77)
say-DIR.3PL.OBV used to FOC my relatives

'... it used **to be said** about my blood relatives ...' (Minde, 1997, 102–103)

(99) ... ômisi mâna **ni-kî-itâ-wak** nitawâsimisak ... (EM 66; .73)
thus used to 1-PST-say-DIR.1SG.3PL my children

'... **I used to tell** my children as follows ...' (Minde, 1997, 36–37)

Three of the five top exemplars, as seen in examples (97), (98), and (99) were forms of *itêw*, 's/he speaks/tells about someone.' In each of the five VTA exemplars, the target verbs were the main and *only* verbs in their clauses. Estimated probabilities ranged form .73 to .91.

ê-Conjunct

(100) 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 .98)
young people

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

(101) ... wâposwa ê-kî-nipahât **ê-wî-kakwê-asam-iko-yâhk** wiya ... (C8GB 13; .97)
rabbits s/he kills him/her CNJ-FUT.VOL-try-feed-INV-3SG.1PL for

'... she killed rabbits and **tried to feed us...'** (Bear et al., 1992/1998, 208–209)

(102) ... “kita-wâpamikot,” “**ê-pê-minihkwât-â-yit**,” itwêw. (VDC2 485–486; .97)
looking at him/her CNJ-come-drink-DIR-3.OBV s/he said

‘... “looking at him,” he said, “**to trade it for a drink**” he said.’ (Vandall & Douquette, 1987, 68–69)

(103) ... 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; .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 & Douquette, 1987, 50–51)

(104) ... miton êsa mân êkotê **ê-kî-isi-sôhkêpit-iko-cik** ... (CMBK-5-2 72; .96)
very apparently used to over there CNJ-PST-thus-support-INV-3OBV.3PL

OFFICIAL TRANSLATION UNAVAILABLE

The majority of the exemplars for the ê-Conjunct Order were the same as those found in the Conjunct outcome in the more general Independent vs. Conjunct alternation, with the notable exception of (104). This is likely due to the fact that the majority of Conjunct forms are, in fact, ê-Conjuncts. Probabilities ranged from .96 to .98.

6.6.3 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

(105) ... namôy êtikwê **ê-miywâsi-k** ôma ta-nipahtâkêhk. (CMBK-5-2 87; .95)
NEG apparently CNJ-be.good-3.SG FOC someone who kills

OFFICIAL TRANSLATION UNAVAILABLE

There was only one correctly identified ê-Conjunct exemplar available, and in this case it was as a main verb of a clause. Its estimated probability was high at .95.

Animate Intransitive Verbs

(106) ... môy tâpwê **ê-ohci-ma-miyomahciho-t** ... (CMBK-4-2 159; .98)
NEG truly CNJ-NEG.PST-RDPLW-feel.well-3.SG

OFFICIAL TRANSLATION UNAVAILABLE

(107) êkos ânima **ê-isi-tâpwê-t** êwako. (JK 160; .97)
so that CNJ-speak.truth-3.SG this

'And he **speaks the truth** in this.' (Kâ-Nîpitêhtêw, 1998, 132–133)

(108) ... ê-wîcêwâyâhk âskaw ê-~ **ê-papâmi-mawiso-t** ... (EM 36–37; .97)
we get along with her sometimes ê-~ CNJ-about-pick.berries-3.SG

'... sometimes going along with her as **she went about berry-picking** ...' (Minde, 1997, 20–21)

(109) **ê-papâmi-pa-pêyako-yân** in the spruce -~- mâka mîn âsay nitâkayâsîmon. (AL 148–149; .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., 1992/1998, 254–255).

(110) êkos ôma nika-mâc-âcimon nîsta, tânisi **ê-isi-ka-kiskisi-yân** ... (CMBK-1-2 14; .97)
so this I will tell bad news I too how CNJ-thus-RDPLW-remeber-1.SG

OFFICIAL TRANSLATION UNAVAILABLE

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 exemplar verbs made use of the discourse preverb {isi-}. Beyond this, the actual semantic criteria of the verbs do not form a cohesive class in this outcome. The estimated probability was high for this outcome, being .97 at the lowest and .98 at the highest.

Transitive Inanimate Verbs

(111) môy wîhkât nânitaw **ê-ohci-itêyihta-mâhk** ... (CMBK-3-2 162; 1.00)
NEG ever simply CNJ-NEG.FAST-think-1PL

OFFICIAL TRANSLATION UNAVAILABLE

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

wîkiwâhk ... (CMBK-4-2 29; .98)
in his/her home

OFFICIAL TRANSLATION UNAVAILABLE

(113) ... tâpiskôc namôya kîkway **ê-itêyihta-hkik** onêhiyâwiniwâw. (VDC2 20–22; .97)
Just like nothing CNJ-think-3.PL their Cree way

'... it is as if their Creeness **means** nothing to them.' (Vandall & Douquette, 1987, 36–37)

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

ê-itwêwiht. (AA 191; .97)
s/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)

(115) ... êkâya kîkway ê-pakitinamâkoyahk, tânisi **ê-itêyiht-am-ahk**. (VDC2 114–115; .90)
nothing s/he allows us how CNJ-think-THM-21.PL

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

Of the top-five VTI exemplars for, only one example, (112), 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 falls under the same umbrella as thinking verbs in the VTI's semantic classes. This is the `TI-nonaction` class. Estimated probabilities were quite high, ranging from .90 to 1.00. Note that in (113), the word *tâpiskôc* is rendered as *tâpiskôt* in print (Vandall & Douquette, 1987, 36–37).

Transitive Animate Verbs

(116) ... ma kîkway wîhkâc **ê-ohci-pakitin-i-cik** aniki nikosis Randy ... (CMBK-2-2 43; .97)
not what ever CNJ-NEG.PST-let.go-INV-3PL.1SG those my son Randy

OFFICIAL TRANSLATION UNAVAILABLE

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

ê-ohci-kiskêyim-ak. (CMBK-4-2 114; .97)
CNJ-NEG.PST-live.with-DIR.1SG.3SG

OFFICIAL TRANSLATION UNAVAILABLE

(118) ... tânsi ê-isi-sîhkimicik, tânisi **ê-isi-nitawêyim-i-cik**, nikî-tôtên. (EM 92; .96)
how s/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)

(119) 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; .93)
my husband over here I come to marry him/her

'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)

(120) ... wiy âh-apisîs piko **ê-kî-asam-ikawi-yâhk**. (CMBK-1-2 25; .92)
for very small a bit CNJ-PST-feed-UNSPEC-1.PL

OFFICIAL TRANSLATION UNAVAILABLE

Although the VTA exemplars for the ê-Conjunct outcome have little cohesion, all but (117) represent a past action, even if not represented in the morphology. Beyond this, ê-Conjunct exemplars often contained first-person goals, as in (116), (118), and (120). These exemplar probabilities ranged from .92 to .96.

kâ-Conjunct

Inanimate Intransitive Verb

(121) ... ita êsa mân êtikwê ê-kî-osâpit, **kâ-kîsikâ-yik** ... (CMBK-5-2 57; .75)
there apparently used to apparently s/he watched from there CNJ-DAY-3.OBV

OFFICIAL TRANSLATION UNAVAILABLE

(122) ... wiya pîhc-âyihk kâ-~ **kâ-pipoh-k** kâ-kî-ayâyâhk ... (C2GB 18; .63)
for inside kâ-~ CNJ-be.winter-3SG we were

'... since we used to be inside **in the winter** ...' (Bear et al., 1992/1998, 70–71)

The kâ-Conjunct outcome for the VII had only two valid exemplars. Each of these exemplars were used not as prototypical verbs, instead referred to temporal periods. Example (121) contains the exemplar verb, a form of *kîsikâw*, which is used as a temporal prepositional phrase/adjunct. In (122) 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, with estimated probabilities not exceeding .75.

Animate Intransitive Verb

(123) ... mîn êkâ awiyak **kâ-kî-minaho-t**, âhci piko pêyakwan
and NEG someone CNJ-PST-hunt-3.SG still a bit one

ê-miyiht wiyâs ... (CMBK-4-2 264; .73)
s/he gives to me meat

OFFICIAL TRANSLATION UNAVAILABLE

(124) ... êkwa **kâ-minaho-cik** ôkik nâpêwak ... (C2GB 14; .72)
and CNJ-hunt-3.PL these men

'... when the men **had killed** an animal ...' (Bear et al., 1992/1998, 68–69)

(125) cikêmô kikî-miyikonaw kôhtâwînaw, kîstanaw **kâ-nêhiyâwi-yahk** ... (JK 7; .63)
of course we were given it by him/her our father we too CNJ-be.Cree-21.PL

'For of course, Our Father has given us, **us who are Crees** ...' (Kâ-Nîpitêhtêw, 1998, 50–51)

(126) ... misatimwak ê-têhtapiyâhk, itê **kâ-minaho-cik** nôhtâwînânak. (CMBK-4-2 250; .59)
horses we ride there CNJ-hunt-3.PL our fathers

OFFICIAL TRANSLATION UNAVAILABLE

(127) êkwa mîna pikw îta **kâ-pîhtikwê-yan** ê-mîcisoyan ... (AL 71; .57)
and also only there CNJ-come.in-2.SG you eat

'And you also ate at every place **you went to** ...' (Bear et al., 1992/1998, 248–249)

Three of the five exemplars, (123), (124), and (126), for the VAIs in the kâ-Conjunct outcome concerned forms of the lexeme *minahow* ('s/he hunts/kills'). Beyond this, there was little that could be generalized about these exemplars. Estimated probabilities were relatively low, ranging from .57 on the low-end to only .73 on the high-end.

Transitive Inanimate Verb

(128) ... 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; .91)

OFFICIAL TRANSLATION UNAVAILABLE

(129) ... 'iyisâhowin' anima ka-~ **kâ-it-am-ihk** aya ... (EM 75; .83)
iyisâhowin this ka-~ CNJ-call-THM-UNSPEC this one

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

(130) 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; .66)
CNJ-want-3.SG

OFFICIAL TRANSLATION UNAVAILABLE

(131) ... anima **kâ-nôhtê-kiskêyiht-ahk** nâha, êwako ê-kî-pawâmit anima ... (SW 39; .61)
that CNJ-want-know-3.SG that one this one s/he had a dream spirit that

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

(132) êwakw ânim ânohc **kâ-mâmiskôt-ahk** ayamihêwiyiniw ... (EM 78; .61)
this that today CNJ-talk.about-3.SG priest ...

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

The VTI exemplars were composed mostly of verbs of speech. The exception to this was *kâ-nôhtê-kiskêyihtahk*, found in (131). This word refers to wanting to know about something. Similar to the VTI exemplars in the ê-Conjunct outcome, those in this outcome all 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 .61 and the highest at .91.

Transitive Animate Verb

(133) ... mâk êkwa awa **kâ-pê-wîhtamaw-it** nitôsim ... (CMBK-4-2 19; .71)
but and this CNJ-come-tell.about-INV.3SG.1SG my stepson

OFFICIAL TRANSLATION UNAVAILABLE

(134) ... kîkway ôki **kâ-wîhtamaw-icik** nitawâsimisak ... (CMBK-4-2 202; .71)
what these CNJ-tell.about-INV.3PL.1SG my children

OFFICIAL TRANSLATION UNAVAILABLE

(135) êwakw âwa **kâ-wîhtamaw-ak** anohc awa, ôta kâ-wîtapimak awa ... (JK 4; .64)
that one that CNJ-tell.about-1SG.3SG today that here I sit with him/her that

'**I told her** this when I was sitting here with her today ...' (Kâ-Nîpitêhtêw, 1998, 46–47)

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 was a kind of speech verb. Probability estimates ranged from .64 to .71.

Other-Conjunct

Intransitive Inanimate Verbs

(136) ... môniyâw ê-pêhtât nêtê **ta-takopay-iyiki** anihi. (CMBK-3-2 134; .53)
white man s/he waits there CNJ-arrive.3.PL.OBV FOC

OFFICIAL TRANSLATION UNAVAILABLE

In the final outcome, the Other-Conjunct, the VII had only a single valid exemplar. Here, the exemplar was a subjunctive form acting as a temporal adjunct. The probability estimate for the one exemplar was quite low at .53.

Animate Intransitive Verbs

(137) ... tânisi k-êtôtamân, **mêstohtê-yêko** pê-miyikawiyâni wêpinâson ... (JK 160; .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)

(138) ... kîspin ê-sôhkêpayik êkotowahk maskihkiy, namôya, namôya ahpô kikaskihtân
if it is strong like that medicine NEG NEG or you are able

ta-nipâ-yan ... (SW 112; .67)
to CNJ-sleep-2.SG

'... if that kind of medicine is very strong, you will not even be able **to sleep** ...' (Whitecalf, 1993, 68–69)

(139) ... êkwa awiyak nôhtê-papâmitâpâsoci, ta-~ **ta-papâmitâpâso-hk** ... (AL 42; .63)
and someone when s/he wants to ride about ta-~ CNJ-ride.around-UNSPEC

'... then if anyone wants to go for a ride ...' (Bear et al., 1992/1998, 244–245)

(140) misawâc ôta, ispî mêht-~ [sic] **mêstohtê-twâwi** ... (JK 18; .53)
in any way here when mêht-~ [sic] die-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 (139), the conditional meaning is still present. Expected probabilities had a large range, from .53 to .89.

Transitive Inanimate Verbs

(141) ... piko kâwi **ka-kîwêtot-am-ahk** k-âtoskêyahk, ka-kakwê-pimâcihoyahk ... (EM 96; .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)

(142) 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; .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)

(143) ... wiya kiyânaw kikaskihtânaw kîkway ka-kî-nipahtamâsoyahk, kayâsi-pimâcihowin
for us we are able to what we killed old way of life

ka-otin-am-ahk ... (C2GB 45; .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., 1992/1998, 82–83)

(144) môy pikw êkosi k-êsi-mâmitonêyihtamahk, **ka-tôt-am-ahk** anima ... (EM 76; .78)
NEG 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)

(145) ... ê-miyohtwât an[a] îskwêw ê-wîcihât anih ôskinîkiwa, **ta-pônihtâ-yit**
s/he is good natured that woman s/he helps him/her that young man, CNJ-quit-3.OBV

minihkwêwin ... (EM 134; .76)
alcohol

'... that woman is good-natured and helps that young man **to quit drinking** ...' (Minde, 1997, 92–93)

The VTI class had five exemplars, all of which occured in the ka/ta-Conjunct. In most cases, these were translated as infinitive forms and nearly always act as non-main verbs. For the VTI exemplars, probability estimates were moderate, ranging from .76 to .85.

Transitive Animate Verbs

(146) ... âta kâ-nisitohtahkik, âta **kitota-twâwi**, tâpiskôc êkâya
although they understand although speak-2PL.3PL.FUT.COND for instance NEG

ê-pêhtâskik ... (VDC2 19–20; .84)
they hear you

'... when in fact they do understand it, it is as if they did not hear you **when you speak to them** ...' (Vandall & Douquette, 1987, 36–37)

(147) k-âyimômâyahk kîc-âyisiyinînaw, ahpô **ka-pâhpih-â-yahk**
when we gossip about him/her our fellow man or laugh-THM.21PL.3PL

ê-kitimâkinâkosit ... (JK 9; .84)
s/he is pitiable

'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, had only two valid exemplars. Both were conditional verbs, though only (146) had the verb in the subjunctive Conjunct form. Instead, (147) 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 (146) the exemplar verb was, as has been seen previously in many instances, one of speech. Both exemplars were well-predicted, with probability estimates of .84.

Chapter 7

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 suggested by the results and how these relate to previous research on Order. Behavioural profiles are 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). 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 in 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.

7.1 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 7.1. Here, each cell is labeled with the outcomes for which an effect is significant. If a cell is coloured green, the effect increased that outcome, while a red cell represents an effect decreasing the likelihood of that outcome. 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 of how the random effect of lexical identity affects Order choice, see section 6.4.4 in Chapter 6.

Table 7.1: Multivariate effects: Independent vs. Conjunct alternation.

	Effects	VII	VAI	VTI	VTA
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`	IND	IND		CNJ
	`PV.discourse`		CNJ	CNJ	
	`PV.position`				CNJ
	`PV.move`				CNJ
Semantic class	`Food`				CNJ
	`Do`			CNJ	
	`Money.count`			CNJ	

Preverbs seemed to only increase the likelihood of a Conjunct form, with two exceptions. 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. This can be done 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 (2014, 140), who suggests the Conjunct to be less likely to be used initially without established context. This description implies the

Conjunct Order is somehow related to the discourse structure of the utterance. The only preverbs associated with the Independent are in the VIIs and 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 VIIs and VAIs and the 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 inflectional 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, `PV.Time`. 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. Proulx (1991) describes this as exemplifying Filmore’s *movement metaphor for time* (1971, 29) wherein time is a continuous process and all things move forward through it. 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-}.

Table 7.2: Multivariate effects: Independent vs. ê-Conjunct.

	Effects	VII	VAI	VTI	VTA
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	

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 those of discourse.

7.2 Independent vs. ê-Conjunct

The pattern of effects is substantially different for the Independent vs. ê-Conjunct alternation than for the Independent vs. Conjunct alternation. This suggests a difference in the *type* of alternation. The effects of this alternation are detailed in Table 7.2. Each cell is labeled with the outcomes for which an effect is significant. If a cell is coloured green, the effect increased that outcome, while a red cell represents an effect decreasing the likelihood of that outcome.

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-morpheme 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 (148), while the ê-Conjunct generally associated with positions on the lower level of the hierarchy (i.e., non-local participants).

(148) $2 > 1 >$ Unspecified Actor $> 3 > 3' > 3''$

As previously discussed, the Independent was generally associated with effects dealing with actors and goals, with the exception of preverbs of time. Again, preverbs of time significantly increased the likelihood of the Independent for both the VIIs and the VAIs and is 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 only for the VTA class. 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.

7.3 Conjunct-Type

The Conjunct-Type alternation was significantly less 'cohesive'. That is, less can be said about an outcome across verb classes. Table 7.3 summarizes the results of this research. Each cell is labeled with the outcomes for which an effect is significant. If a cell is coloured green, the effect increased that outcome, while a red cell represents an effect decreasing the likelihood of that outcome. As can be seen, even when an effect is present in multiple verb classes, it is not always the case that the effect significantly increased or significantly decreased the likelihood of the same outcome in each inflectional class. For example, while `PV.Discourse` is significant for VAIs and VTAs, the effect increases the likelihood of ê-Conjunct in both. Conversely, it significantly decreases the likelihood of kâ-Conjunct in the VAIs and Other-Conjunct in the VTAs. Similarly, `PV.Position` significantly increased the likelihood of the ê-Conjunct in the VAIs, VTIs, and VTAs. Conversly, It significantly decreases the likelihood of kâ-Conjunct in VAIs and VTAs while also decreasing the likelihood of the Other-Conjunct in the VTIs.

Table 7.3: Multivariate effects: Conjunct-Type alternation.

		VII	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 except 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 actor morphemes while the Other-Conjunct was increased for the same. 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 of the 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 (149), where the verb *kâ-pipohk* is used to mean 'in winter,' rather than being used as a more declarative statement.

(149) ... awâsisak wâwîs **kâ-pipoh-k**.
children especially CNJ-be.winter-3SG

'... especially for children **in winter**.' (Minde, 1997, 137)

7.4 Model Statistics

Modelling produced varied results. With the main mixed-effects models in the Independent vs. Conjunct alternation, only the VAI and VTA models were quantitatively well-fitting. This may simply be due to frequency effects. The Independent vs. ê-Conjunct alternation showed more success in model fitting: only the VIIs failed to produce a relatively well-fitting 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-fitting models. Assuming this pattern is not due to data-specific effects, it may be the case 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-effects and semantic-effects).

A number of previous studies 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-effects 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 highlight the importance of semantic classification of subjects and infinitives when choosing between verbs of *try* in Russian; and Abdulrahim

(2013), who found that semantic classifications can be useful in profiling *go* and *come* verbs in Arabic. These results can be compared with those of Klavan (2012), who 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 inflectional class, overall the difference between semantic-effects and morphological-effects was minimum for the Nêhiyawêwin Order alternation.

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, as they lack a wide system of preverbs or polypersonal agreement, for example. 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 investigated in the above studies is different than those in this dissertation. While Arppe (2008), Divjak and Arppe (2013), and Abdulrahim (2013) focused on an alternation between near-synonymous terms to do generally with a single semantic domain (*thinking*, *trying*, and *motion* verbs respectively), the alternation of Order is more structural, as described in Chapter 2. Only the analysis of Kalvan (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 use 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 important.

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.

It is also important to keep in mind that results can be affected by the make up of the corpus being analyzed. Although the corpus, described in Chapter 3, contains a variety of speakers and target domains, it was a collection of mostly elderly speakers. The types of topics and perhaps even constructions used might be different than a younger set of speakers. These differences could result in different findings, should similar research be undertaken with a corpus of different speakers.

7.5 Cross-Validation

Cross-Validation, at least based on the accuracy statistic, suggested that modelling in most cases was generalizable. The only exception to this finding was in the the VII model for the Independent vs. Conjunct alternation, where the accuracy of the mixed-effects model of this dissertation was 77%, as compared to the 81%–82% 'range' of the cross-validation models. Further, the cross-validation models had a mean and median of 81% for accuracy. Interestingly, it is the main mixed-effects model that performs with worse accuracy in this case. Given data it has not seen, the VII model that is trained with the same effects performs better on unseen data than on the data it was trained with. This pattern is essentially the opposite of overfitting, and suggests that the main mixed-effects model is failing to learn the important features affecting the choice between the outcomes (Cunningham & Delany, 2021). This result may be explained by the fact that only lemma identity and `PV.time` were used as predictors in this model; however, this was also the case for the Independent vs. ê-Conjunct alternation for the VII model, where the main mixed-effects model was within the accuracy range of the cross-validation sets. The exact reason for this under-fitting and the difference between the two alternations is unclear, but it may be that `PV.time` was simply more important in determining Order in the Independent vs. Conjunct alternation. It is also worth noting that the VII cross-validation models had a large accuracy range (10 percentage points) in the Independent vs. ê-Conjunct case as compared to the two point 'range' in the Independent vs. Conjunct alternation.

For a more in-depth discussion on cross-validation in this dissertation, and for a consideration of other model statistics, see Appendix B.

7.6 Exemplar Extraction

Exemplar extraction was mostly successful, and useful exemplars were presented in Chapter 6. Interestingly, in clustering tokens to extract exemplars, a pattern emerged. 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 lexeme is a full 10 percentage points lower). By choosing to ignore all but one *itwêw* lemma, one would be highlighting less-certain predictions. Further, VAI quotatives are closely related to the Independent, this may suggest that the Independent is construction that is more concerned with basic statement of facts, rather than more complex discourse structures. Further analysis of quotatives specifically is needed.

7.7 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 (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-fitting models of the Conjunct-Type alternation, there was little in terms of clear pattern that helped create a linguistically informative explanation of the phenomenon. In nearly all cases, no predictor was significant for more than half of the inflectional classes. There were two exceptions to this. First, the `PV.Position` predictor which was significant for the VAIs, VTIs, and VTAs and always increased the likelihood of an ê-Conjunct. Second, the `Actor.2` was significant for the VAIs, VTIs, and

[1]This preverb was included in the `PV.Quantity` class due to its straightforwardly quantifier-like semantics.

VTAs, 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...' That there was less linguistic informativity may result from the fact that the Conjunct types are much more clearly associated with certain semantic interpretations (e.g., the subjunctive Conjunct always produces a hypothetical statement), and it is this higher-level linguistic information that is driving the alternation, rather than the other features assessed in this dissertation.

Finally, it appears that, although mixed-effects models were always the best-fitted models overall, individual inflectional classes sometimes showed other models out-performing the mixed-effects models. This is exemplified (but does not cover other alternations) in Table 6.49, reproduced as 7.4 here, where the VII class exhibited higher ρ^2 in LEM models than in the mixed-effects case.

Table 7.4: Model comparisons for the Independent vs. Conjunct alternation. Green cells with bold items represent a very good model fit, per McFadden (1973). (repeated from page 136)

	VII		VAI		VTI		VTA	
	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2	τ_c	ρ^2
SEM.MORPH	.31	.05	.47	.15	.32	.09	.35	.10
LEM	.36	.12	.52	**.26**	.34	.15	.42	.18
SEM.LEM	.35	.12	.52	**.27**	.36	.15	.44	.19
MORPH.LEM			.52	**.27**	.39	.16	.43	.19
ME	.35	.12	.53	**.27**	.42	.16	.45	**.21**
SEM	.31	.05	.47	.14	.30	.06	.39	.08
MORPH			.26	.03	.28	.04	.24	.04

Although the Lemma-as-random-effect models were rarely well-fitting enough or possessing substantial variation explanation, they still often performed well. In Table 7.4, τ scores are all over .30 in the lemma-only model, and ρ^2 values ranged from .12 to .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. Much of the alternation appears to be lexically motivated.

Chapter 8

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 alternate outcomes; that is, it predicts the proportion of how likely alternative outcomes are to occur based on some context on the long run.

As previously mentioned, the main research question of this dissertation has been:

1. What morphosyntactic and semantic features affect a lemma's propensity to occur in a particular alternation of Order?

Although the variables predicting Order varied by inflectional class and alternation type, speech-actor arguments and preverbs indicating a non-present tense increased the likelihood of an Independent. As well, preverbs indicating some sense of discourse-level relationship increased the likelihood of a Conjunct form.

Following from this main question, a secondary research question was presented along with two predictions:

2. Can Order choice be sufficiently predicted by primarily morphosyntactic and semantic predictors?

(a) Because Nêhiyawêwin is a morphologically rich language, and also due to the findings of previous similar alternation studies (Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013), the morphosyntactic and semantic predictors will provide substantial explanation of variation in modelling the alternations, though some variation will remain due to a lack of syntactic information (e.g., argument structure).

(b) Semantic classification of constituents will do more to predict all alternations than morphosyntactic variables (as in Abdulrahim, 2013; Arppe, 2008; Divjak & Arppe, 2013).

In regards to research question (2), the results were mixed. Prediction (2a) was somewhat supported: models often, but not always, exceeds the ρ^2 threshold of .20 that identifies a well-fitted model using semantic and morphosyntactic predictors. This pattern indicates that purely syntactic information was not the only way to investigate the phenomenon. Preverbs indicating discourse-level relationships often proved significant predictors, though their status as morphological vs. syntactic could be disputed. While morphological predictors appeared to be helpful in predicting Order, they were not able to fully explain the phenomenon. That not all models produced well-fitting models suggests that the predictors used in this dissertation were not sufficient to entirely explain these alternations.

Prediction (2b) was ultimately 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 lemma's semantic class might affect Order choice.

While investigating the two research questions of this dissertation, an interesting pattern appeared: Although all modelling provided some insight, some models were more linguistically informative than others. Despite not producing the models that were most well-fitting, the Independent vs. ê-Conjunct models presented clear motivations. In this alternation, speech-act participants increased the likelihood of the Independent Order and discourse-level preverbs increased the likelihood of the ê-Conjunct. This straightforwardness in motivation was also seen in the Independent vs. Conjunct alternation. The Conjunct-Type alternation had the best-fitted models, with all inflectional classes' models being relatively well-fitting. On the other hand, in the Independent vs. Conjunct alternation only the VAI and VTA models were above the threshold for a good fit. For the Independent vs. ê-Conjunct alternation, only the VAI, VTI, and VTA classes had ρ^2 scores above the level that suggests a well-fitted model. Despite the well-fitting models, the actual linguistic informativity of the models from the Conjunct-Type alternation did not present a cohesive set of motivations. 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 of context. Given that the ê-Conjunct was simply so much more numerous than other outcomes, this proved a valid strategy. It is important to keep in mind that the concept of *linguistic informativity* or even that of a 'straightforward narrative' emerging from the results of a study are, to some extent, influenced by the researcher. As a non-native speaker who lacks the cultural perspective that a member of the Nêhiyawêwin language community might have, it is possible that some patterns of interpretation of the results were not obvious despite careful analysis. At the same time, an outsider's perspective can allow for identification of phenomena that native speakers may be 'blind' to, due to their familiarity with the language. Future research would benefit from co-investigation with a native speaker of Nêhiyawêwin who is also knowledgeable about the linguistics of the language.

In addressing these research questions and predictions, 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-effects modelling to help understand the phenomena. Specifically, this research confirmed that, when deciding between an 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 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. This focus would likely produce much better modelling, especially if combined with the morphological and semantic 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. Alternative techniques other than logistic regression[1] could be used, including those techniques mentioned in Section 5.4. Further, if much of this alternation rests in higher-level phenomena and contextual use in the discourse, and if context is something

[1]One could also consider combining logistic regression with neural technologies to improve results. This sort of synergy is seen in Tzougas and Kutzkov (2023), who document 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. To my knowledge, this technique is not widely used yet, and so its efficacy is not yet clear.

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.[2] As Arppe and Järvikivi (2007) and Klavan (2012) 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. A major motivation for this dissertation was to inspire hypotheses that researchers can use to best target their research. In this way, participants can take part in only those tasks that are most likely to produce informative results. Given that the majority of fluent Nêhiyawêwin speakers are elderly, the amount of time participants can make available is usually substantially lower than in other participant groups.

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 statistical analysis techniques, 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

[2] I, along with Katherine Schmirler, 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 a form of conducting such an experiment that would be more agreeable to the speakers could be developed.

quantitative and linguistic analyses in this dissertation may be esoteric to students of Nêhiyawêwin without prior knowledge of these topics, it is my hope that the exemplars provided in Chapter 6 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, at least in some cases the choice of Order in Nêhiyawêwin is a gradient, rather than definite, choice.

References

Abdulrahim, D. (2013). *A corpus study of basic motion verbs in modern standard arabic*. (Doctoral Dissertation, University of Alberta) doi: https://doi.org/10.7939/R35T12

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). John Wiley & Sons.

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* (Wolfart, H. C. & F. Ahenakew, Eds.). University of Manitoba Press.

Alcázar, A., & Saltarelli, M. (2014). *The syntax of imperative clauses: A performative hypothesis*. Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511794391.004

Arppe, A. (2008). *Univariate, bivariate, and multivariate methods in corpus-based lexicography—A study of synonymy*. University of Helsinki. Retrieved from `https://helda.helsinki.fi/handle/10138/19274` (Doctoral Dissertation, University of Helsinki. Publications of the Department of General Linguistics, No. 44)

Arppe, A. (2009). Linguistic choices vs. probabilities—how much and what can linguistic theory explain? In S. Featherston & S. Winkler (Eds.), *The fruits of empirical linguistics* (Vol. 1, pp. 1–24). De Gruyter Mouton. doi: https://doi.org/10.1515/9783110216141.1

Arppe, A. (2013). Polytomous: Polytomous logistic regression for fixed and mixed effects [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=polytomous` (R package version 0.1.6)

Arppe, A., & Järvikivi, J. (2007). Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, *3*(2), 131–159. doi: https://doi.org/10.1515/CLLT.2007.009

Arppe, A., Schmirler, K., Harrigan, A. G., & Wolvengrey, A. (2020). A morphosyntactically tagged corpus for plains cree. In M. Macaulay & M. Noodin (Eds.), *Papers of the forty-ninth Algonquian conference* (pp. 1–16). Michigan State University Press. doi: https://doi.org/10.14321/J.CTVV417GP.5

Baayen, R. H. (2012). Mixed-effects models. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 668–677). Oxford University Press. doi: https://doi.org/10.1093/OXFORDHB/9780199575039.013.0022

Barth, D., & Kapatsinski, V. (2018). Evaluating logistic mixed-effects models of corpus- linguistic data in light of lexical diffusion. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-effects regression models in linguistics* (pp. 99–116). Springer. doi: https://doi.org/10.1007/978-3-319-69830-4_6

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bear, G., Calliou, I., Fritz, J., Fraser, M., Lafond, A., & Wells, R. L. M. (1998). *kôhkominawak otâcimowiniwâwa/Our grandmothers' lives: As told in their own words* (F. Ahenakew & H. C. Wolfart, Eds.). Canadian Plains Research Centre. (Original work published 1992)

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bloomfield, L. (1930). *Sacred stories of the Sweet Grass Cree.* F.A. Acland, Printer to the King's Most Excellent Majesty.

Bloomfield, L. (1946). Algonquian. In H. Hoijer et al. (Eds.), *Linguistic structures of Native America* (pp. 85–129). Viking Fund Publications in Anthropology. Retrieved 2024-02-06, from https://home.cc.umanitoba.ca/~oxfordwr/bloomfield1946/

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Royal Netherlands Academy of Arts and Sciences.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, *10*(4), 417–451. doi: https://doi.org/10.2307/3001616

Cook, C. (2014). *The clause-typing system of Plains Cree: Indexicality, anaphoricity, and contrast.* Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780199654536.001.0001

Cook, C., & Muehlbauer, J. (2010). *A morpheme index of Plains Cree. Unpublished manuscript.* Retrieved 2023-12-31, from https://www.academia.edu/304874/A_morpheme_index_of_Plains_Cree

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review*, *30*(3), 341–353. doi: https://doi.org/10.2307/2090715

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton University Press.

Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics.* Oxford University Press.

Cunningham, P., & Delany, S. J. (2021). Underestimation bias and underfitting in machine learning. In F. Heintz, M. Milano, & B. O'Sullivan (Eds.), *Trustworthy AI–integrating learning, optimization and reasoning* (pp. 20–31). Springer.

Dacanay, D., Arppe, A., & 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 A. Arppe, J. Good, A. Harrigan, J. L. Mans Hulden, A. P. Sarah Moeller, & L. S. Miikka Silfverberg (Eds.), *Proceedings of the 4th workshop on the use of computational methods in the study of endangered languages* (Vol. 1, pp. 33–43). Association for Computational Linguistics.

D'Agostino, R. B., Chase, W., & 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. doi: https://doi.org/10.2307/2685002

Dahlstrom, A. (2014). *Plains Cree morphosyntax* (2nd ed.). Routledge. doi: https://doi.org/10.4324/9781315852225

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies, NAACL-HLT* (Vol. 1, pp. 4171–4186). Association for Computational Linguistics. doi: 10.18653/V1/N19-1423

DiMarco, C., Hirst, G., & Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI spring symposium on building lexicons for machine translation* (Vol. 1, pp. 114–121).

Divjak, D. (2010). *Structuring the lexicon: A clustered model for near-synonymy*. Walter de Gruyter. doi: https://doi.org/10.1515/9783110220599

Divjak, D., & Arppe, A. (2013). Extracting prototypes from exemplars what can corpus data tell us about concept representation? *Cognitive Linguistics*, *24*(2), 221–274. doi: https://doi.org/10.1515/COG-2013-0008

Divjak, D., & Gries, S. T. (2006). Ways of trying in russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, *2*(1), 23–60. doi: https://doi.org/10.1515/CLLT.2006.002

Domencich, T. A., & McFadden, D. (1975). *Urban travel demand—A behavioral analysis*. North-Holland Publishing. Retrieved 2024-02-06, from `https://eml.berkeley.edu/~mcfadden/travel.html`

Eberhard, D., Simons, G. F., & Fennig, C. D. (Eds.). (2023). *Ethnologue: Languages of the world* (26th ed.). SIL International. (Online version: `https://www.ethnologue.com/`)

Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, *28*(2), 105–144. doi: https://doi.org/10.1162/089120102760173625

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness.* Cambridge University Press. doi: https://doi.org/10.1017/9781316584880

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyad (Eds.), *Proceedings of second international conference on knowledge discovery and data mining* (p. 226–231). Portland, Oregon, United States of America: AAAI Press.

Fellbaum, C. (1998). *Wordnet: An electronic lexical database.* MIT press. doi: https://doi.org/10.7551/MITPRESS/7287.001.0001

Filmore, C. J. (1971). *Lectures on deixis 1971.* Indiana University Linguistics Club.

Firth, J. R. (1962). A Synopsis of Linguistic Theory 1930–1955. In *Studies in linguistic analysis* (pp. 1–32). Basil Blackwell. (Special volume of the Philological Society)

Fox, J. (2016). *Applied regression analysis, linear models, and related methods.* (3rd ed.). SAGE Publications.

Frank, E., & Kramer, S. (2004). Ensembles of nested dichotomies for multi-class problems. In *Proceedings of the twenty-first international conference on machine learning* (pp. 305–312). Association for Computing Machinery. doi: https://doi.org/10.1145/1015330.1015363

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764. doi: https://doi.org/10.2307/2281536

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* Continuum.

Gries, S. T., & Divjak, D. (2009). Behavioral profiles: a corpus-based approach to cognitive semantic analysis. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57–75). John Benjamins B. V.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Gries, S. T., & Stoll, S. (2009). Finding developmental groups in acquisition data: Variability-based neighbour clustering. *Journal of Quantitative Linguistics*, *16*(3), 217–242. doi: https://doi.org/10.1080/09296170902975692

Grundt, A. W. (1978). The functional role of the Indo-European theme vowel. *Pacific Coast Philology*, *13*, 29–35. doi: https://doi.org/10.2307/1316361

Hamp, B., & Feldweg, H. (1997). Germanet—A lexical-semantic net for German. In P. Vossen, G. Adriaens, N. Calzolari, A. Sanfilippo, & Y. Wilks (Eds.), *Automatic information extraction and building of lexical semantic resources for nlp applications.* Association for Computational Linguistics.

Han, W., Arppe, A., & Newman, J. (2013). Topic marking in a shanghainese corpus: from observation to prediction. *Corpus Linguistics and Linguistic Theory*, *13*(2), 1–29. doi: https://doi.org/10.1515/CLLT-2013-0014

Hand, D., & Christen, P. (2017). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, *28*(3), 539–547. doi: 10.1007/s11222--017--9746--6

Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, *1*(1), 75–98. doi: https://doi.org/10.1075/IJCL.1.1.06HAN

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MIT Press. doi: https://doi.org/10.7551/MITPRESS/9780262018579.001.0001

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer. doi: https://doi.org/10.1007/978-3-319-19425-7

Harrigan, A. G., & Arppe, A. (2015, October 22–23). *Oswêw êkwa ê-tipiskâk, mâka mâcîw âhpô ê-mâcît?* [Conference presentation]. 47th Algonquian Conference, Winnipeg, Manitoba, Canada. Unpublished.

Harrigan, A. G., & Arppe, A. (2021). Leveraging english word embeddings for semi-automatic semantic classification in Nêhiyawêwin (Plains Cree). In M. Mager et al. (Eds.), *Proceedings of the first workshop on natural language processing for Indigenous languages of the Americas* (pp. 113–121). Association for Computational Linguistics. doi: https://doi.org/10.18653/V1/2021.AMERICASNLP-1.12

Harrigan, A. G., & Arppe, A. (2023). Leveraging Majority Language Resources for Plains Cree Semantic Classification. In M. Macaulay & M. Noodin (Eds.), *Papers of the fifty-second Algonquian conference* (pp. 129–146). Michigan State University Press. doi: https://doi.org/10.14321/j.ctv32r03jv.7

Harrigan, A. G., & Arppe, A. (in press). Plains Cree Order as alternation. *Linguistics Vanguard*.

Harrigan, A. G., Arppe, A., & Wolvengrey, A. (2018). Toward a Detailed Plains Cree VAI Paradigm. In M. Macaulay & M. Noodin (Eds.), *Papers of the forty-seventh Algonquian conference* (pp. 129–146). Michigan State University Press. doi: https:/doi.org/10.14321/J.CTT1X76DJ5.10

Harrigan, A. G., Schmirler, K., Arppe, A., Antonsen, L., Trosterud, T., & Wolvengrey, A. (2017). Learning from the computational modelling of Plains Cree verbs. *Morphology*, *27*(4), 565–598. doi: https://doi.org/10.1007/S11525-017-9315-X

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146–162. doi: https://doi.org/10.1080/00437956.1954.11659520

Hirst, G. (1997, November 8–10). Context as a spurious concept [Presentation]. AAAI Fall Symposium on Context in Knowledge Representation and Natural Language, Cambridge, United States of America. Unpublished invited talk. doi: https://doi.org/10.48550/ARXIV.CMP-LG/9712003

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. doi: https://doi.org/10.1002/0471722146

Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics*, *32*(2), 223–262. doi: https://doi.org/10.1162/COLI.2006.32.2.223

Jacques, G., & Antonov, A. (2014). Direct/inverse systems. *Language and Linguistics Compass*, *8*(7), 301–318. doi: https://doi.org/10.1111/LNC3.12079

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer. doi: https://doi.org/10.1007/978-1-4614-7138-7

Jolley, C. (1983). Algonquian person hierarchy: Morphosyntactic or semantic? In W. Cowan (Ed.), *Actes du quatorziéme congrès des Algonquinistes* (pp. 281–292). Ottawa: Carleton University.

Kâ-Nîpitêhtêw, J. (1998). *ana kā-pimwēwēhahk okakēskihkēmowina/The counselling speeches of Jim Kâ-nîpitêhtêw* (F. Ahenakew & H. C. Wolfart, Eds.). University of Manitoba Press.

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and visualize the results of multivariate data analyses.* (*1.0.7*). Retrieved from https://CRAN.R-project.org/package=factoextra

Klavan, J. (2012). *Evidence in linguistics: Corpus-linguistic and experimental methods for studying grammatical synonymy*. University of Tartu Press. Retrieved from http://hdl.handle.net/10062/27865 (Doctoral Dissertation, University of Tartu. Dissertationes Linguisticae Universitatis Tartuensis No. 15)

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. doi: https://doi.org/10.1515/CLLT-2016-0010

Klavan, J., & Divjak, D. (2016). The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica*, *50*(2), 355–384. doi: https://doi.org/10.1515/FLIN-2016-0014

Klecka, R., William. (1980). *Discriminant analysis.* SAGE Publications. doi: https://doi.org/10.4135/9781412983938

König, E., & Siemund, P. (2007). Speech act distinctions in grammar. In T. Shopen (Ed.), *Language typology and syntactic description* (2nd ed., Vol. 1, pp. 276–324). doi: https://doi.org/10.1017/CBO9780511619427.005

Kviz, F. J. (1981). Interpreting Proportional Reduction in Error Measures as Percentage of Variation Explained. *The Sociological Quarterly*, *22*(3), 413–420. doi: https://doi.org/10.1111/J.1533-8525.1981.TB00671.X

Lakoff, G. (1984). Performative subordinate clauses. In C. Brugman, M. Macauly, A. Dahlstrom, M. Emanatian, & B. Moonwomon (Eds.), *Proceedings of the tenth annual meeting of the Berkley Linguistics Society.* Berkley Linguistics Society. doi: https://doi.org/10.3765/BLS.V10I0.1953

Landauer, T., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. doi: https://doi.org/10.1080/01638539809545028

LeClaire, N., Cardinal, G., & Hunter, E. (1998). *Alberta elders' cree dictionary/alperta ohci kehtehayak nehiyaw otwestamâkewasinahikan* (E. Waugh, Ed.). University of Alberta Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Lewis, P. M., & Simons, G. F. (2012). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, *55*(2), 103–120.

Li, L., & Sporleder, C. (2010). Using Gaussian mixture models to detect figurative language in context. In R. Kaplan, J. Burstein, M. Harper, & G. Penn (Eds.), *Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics* (pp. 297–300). Association for Computational Linguistics.

Liebetrau, A. M. (1983). *Measures of association*. SAGE Publications. doi: https://doi.org/10.4135/9781412984942

Majerova, I., & Nevima, J. (2017). The measurement of human development using the ward method of cluster analysis. *Journal of International Studies*, *10*(2). doi: https://doi.org/10.14254/2071-8330.2017/10-2/17

Majewska, O., Vulić, I., McCarthy, D., & Korhonen, A. (2020). Manual clustering and spatial arrangement of verbs for multilingual evaluation and typology analysis. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th international conference on computational linguistics* (pp. 4810–4824). International Committee on Computational Linguistics. doi: https://doi.org/10.18653/V1/2020.COLING-MAIN.423

Masuskapoe, C. (2010). *piko kîkway ê-nakacihtât: kêkêk otâcimowina ê-nêhiyawastêki/There's nothing she can't do: Kêkek's autobiography published in Cree* (H. C. Wolfart & F. Ahenakew, Eds.). Algonquian and Iroquoian Linguistics.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1977). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments. *Cowles Foundation Discussion Papers* (474). Retrieved 2024-02-06, from `https://elischolar.library.yale.edu/cowles-discussion-paper-series/707`

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). SAGE Publications. doi: https://doi.org/10.4135/9781412983433

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio & Y. LeCun (Eds.), *1st international conference on learning representations: Workshop track proceedings.* doi: https://doi.org/10.48550/ARXIV.1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Proceedings of the 26th international conference on neural information processing systems* (Vol. 2, pp. 3111–3119). Curran Associates. doi: https://doi.org/10.48550/ARXIV.1310.4546

Minde, E. (1997). *kwayask ê-kî-pê-kiskinowâpahtihicik/Their example showed me the way: A Cree woman's life shaped by two cultures* (F. Ahenakew & H.C. Wolfart, Eds.). University of Alberta Press.

Okimāsis, J. L. (2018). *Cree: Language of the Plains/nēhiyawēwin: paskwāwi-pīkiskwēwin* (3rd ed.). 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. doi: https://doi.org/10.1080/14786440009463897

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (*EMNLP*) (pp. 1532–1543). Association for Computational Linguistics. doi: https://doi.org/10.3115/V1/D14-1162

Pijpops, D. (2020). What is an alternation?: Six answers. *Belgian Journal of Linguistics*, *34*(1), 283–294. doi: https://doi.org/10.1075/BJL.00053.PIJ

Proulx, P. (1991). Proto-Algic III: Pronouns. *Kansas Working Papers in Linguistics*, *16*, 129–170. doi: https://doi.org/10.17161/KWPL.1808.429

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ratt, S. (2016). *Mâci-nêhiyawêwin/beginning cree*. University of Regina Press. doi: https://doi.org/10.1515/9780889774360

Sadock, J. M., & Zwicky, A. M. (1985). Speech act distinctions in syntax. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 1). Cambridge University Press.

Särndal, C. (1974). A comparative study of association measures. *Psychometrika*, *39*, 165–187. doi: https://doi-org/10.1007/BF02291467

Schmirler, K. (2022). *Syntactic features and text types in 20th century plains cree: A constraint grammar approach*. (Doctoral Dissertation, University of Alberta) doi: https://doi.org/10.7939/r3-pz87-ye25

Schmirler, K., Arppe, A., Trosterud, T., & Antonsen, L. (2018). Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments. In N. Calzolari et al. (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation* (*LREC 2018*) (pp. 2981–2988). European Language Resources Association (ELRA).

Searle, J. R., & Vanderveken, D. (1985). Speech acts and illocutionary logic. In D. Vanderveken (Ed.), *Logic, thought and action* (pp. 109–132). Springer. doi: https://doi.org/10.1007/1-4020-3167-X_5

Snoek, C., Thunder, D., Loo, K., Arppe, A., Lachler, J., Moshagen, S., & Trosterud, T. (2014). Modeling the noun morphology of Plains Cree. In J. Good, J. Hirschberg, & O. Rambow (Eds.), *Proceedings of the 2014 workshop on the use of computational methods in the study of endangered languages* (pp. 34–42). Association for Computational Linguistics. doi: https://doi.org/10.3115/V1/W14-2205

Statistics Canada. (2023). *Census Profile.* 2021 Census of Population. Statistics Canada Catalogue no. 98-316-X2021001. (Table). Retrieved from `https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E` (Knowledge of languages)

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, *2*(1), 23–55. doi: https://doi.org/10.1075/fol.2.1.03stu

Takahashi, H. (2008). Imperatives in concessive clauses: compatibility between constructions. *Constructions*, *5*. doi: https://doi.org/10.24338/CONS-454

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, *76*(1), 103–154. doi: https://doi.org/10.1086/224909

Tzougas, G., & Kutzkov, K. (2023). Enhancing logistic regression using neural networks for classification in actuarial learning. *Algorithms*, *16*(2). doi: https://doi.org/10.3390/A16020099

Vandall, P., & Douquette, J. (1987). *wâskahikaniwiyiniw-âcimowina/Stories of the house people* (F. Ahenakew, Ed.). University of Manitoba Press.

Visitor, L., Junker, M.-O., & Neacappo, M. (Eds.). (2013). *Eastern James Bay Cree thematic dictionary* (*southern dialect*). Cree School Board.

Wang, S., & Bond, F. (2013). Building the Chinese Open Wordnet (COW): Starting from Core Synsets. In P. Bhattacharyya & K.-S. Choi (Eds.), *Proceedings of the 11th workshop on Asian language resources* (pp. 10–18). Asian Federation of Natural Language Processing.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. doi: https://doi.org/10.1080/01621459.1963.10500845

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. doi: https://doi.org/10.1080/00031305.2016.1154108

Whitecalf, S. (1993). *kinêhiyâwiwininaw nêhiyawêwin/The Cree language is our identity: The La Ronge lectures of Sarah Whitecalf* (H. C. Wolfart & F. Ahenakew, Eds.). University of Manitoba Press.

Wolfart, H. C. (1973). Plains Cree: A grammatical study. *Transactions of the American Philosophical Society*, *63*(5). doi: https://doi.org/10.2307/1006246

Wolfart, H. C. (1996). Sketch of Cree, an Algonquian language. In I. Goddard (Ed.), *Handbook of North American Indians: Languages* (Vol. 17, pp. 390–439). Smithsonian Institution.

Wolvengrey, A. (2001). *nêhiyawêwin itwêwina/Cree: Words* (Bilingual ed.). Regina: University of Regina Press.

Wolvengrey, A. (2011). *Semantic and pragmatic functions in Plains Cree syntax*. LOT. Retrieved from `https://hdl.handle.net/11245/1.342704` (Doctoral Dissertation, Universiteit van Amsterdam. LOT dissertation series No. 268)

Wolvengrey, A. (2012, June 6–8). The verbal morphosyntax of aspect-tense-modality in dialects of Cree [Presentation]. 2nd International Conference on Functional Discourse Grammar, Ghent, Belgium. Unpublished talk.

Wu, J. (2012). *Advances in k-means clustering: a data mining thinking.* Springer. doi: https://doi.org/10.1007/978-3-642-29807-3

Yu, L.-C., Wang, J., Lai, K. R., & Zhang, X. (2017). Refining word embeddings for sentiment analysis. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 534–539). Association for Computational Linguistics. doi: https://doi.org/10.18653/V1/D17-1056

Appendix A

Corpus Fact Sheet

A.1 Created

- Compiled for electronic corpus 2015-2020
- Collected for initial recording: 1970s-1990s

A.2 Research Goal

"Morphosyntactically tagged corpora contribute to linguistic descriptions and language maintenance in a variety of ways. Corpora can be used to supplement online dictionaries with examples of forms in natural language use and allow for systematic quantitative analyses to be performed on much larger scales than previously possible, without extensive experience in computational techniques. Such analyses can then further inform qualitative analysis and benefit descriptions overall." Arppe et al. (2020, 1)

A.3 Documentation

- Arppe et al. (2020)

- Schmirler (2022, 122-123)

A.4 Domain Considerations

Target Domains:

- Fluent
- Nêhiyawêwin
- General language
- Conversation
- Oration

A.5 Operational Domain

- Full list of sources
- No explicit operational domain
- Resources are mostly available through academic databases, though some are not. Translations are not freely available in all cases. Full text is almost entirely unavailable.

A.6 Strata

- Not Stratified

A.7 Registers

- Eight registers: Sacred stories, old-time stories, counseling stories, funny stories, and personal stories.
- Exact Distributions unclear

A.8 Sampling unit

- Full texts

A.9 Sampling methods

- Nonrandom, convenience sample

A.10 Evaluation

- No evaluation

A.10.1 Linguistic variable

- N/A

A.11 Sample size

Texts:

- 17
- Tokens: 142,192; 80,221 in Nêhiyawêwin

A.12 Evaluation

- No empirical evaluation

Appendix B

Cross-Validation

As mentioned in Chapter 7, a 10-fold cross-validation was undertaken for the main mixed-effects models presented in this dissertation to judge how generalizable the models were. To perform cross-validation, k-fold validation as described in James et al. (2013, 181) was chosen as a method.

While Chapter 7 only gave cross-validation results for accuracy, this appendix provides details on the ρ^2 and τ_c measures are also given. As in the main body of this dissertation, where ME results fell within the cross-validation (CV) range, this was considered evidence that the models were reasonably in line with the cross-validation set (and thus reasonably justifiable). Where an ME measure fell within the CV range, the cell was reported normally. Where an ME measure did not fall within the CV range, but it differed by .03 or less than the CV mean *or* median (or three percentage points or lower in the case of accuracy), then it was reported in an orange cell. Orange cells represent a 'nearly-validated' measure. Where an ME measure did not fall within the CV range, *and* it differed by more than .03 (or three percentage points for accuracy), then it was reported in a red cell. Red cells represent a measure who's generalizability can be called into question.

B.0.1 Independent vs. Conjunct

Table B.1: Cross-Validation of the Independent vs. Conjunct models.

	CV Min	CV Max	CV Mean	CV Median	ME
VII					
Accuracy	81%	82%	81%	81%	77%
R2.likelihood	.06	.13	.11	.11	.13
tau.classification	.35	.39	.36	.36	.35
VAI					
Accuracy	77%	78%	78%	78%	79%
R2.likelihood	.19	.26	.25	.25	.27
tau.classification	.49	.51	.51	.51	.53
VTI					
Accuracy	72%	78%	77%	77%	76%
R2.likelihood	.08	.14	.13	.13	.16
tau.classification	.30	.46	.42	.43	.42
VTA					
Accuracy	72%	74%	73%	73%	75%
R2.likelihood	.14	.21	.19	.20	.21
tau.classification	.42	.46	.44	.44	.45

Cross-Validation results, given in Table B.1, showed that the mixed-effects (ME) modelling for the Independent vs. Conjunct alternation was relatively generalizable. Although accuracies for the VII, VAI, and VTA ME models were out of the CV range, only the VII showed a large difference. The VAI and VTA ME models both only deviated from the CV mean/median by 1 percentage point. The accuracy for the VTI ME model was within the CV range and only slightly less than the mean/median.

The ρ^2 score for the VII and VTAs were both within the CV range and each .02 higher than the mean. The VAI ME model's ρ^2 was just outside of the CV range and .02 higher than the CV mean/median, with the VTI model's ρ^2 showing a similar pattern being .02 higher than the CV maximum and .03 higher than the CV mean/median.

The τ_c scores for the VII, VTI, and VTA ME models were all within the CV range and only different from their respective CV means by .01 at most. The VAI ME model showed a slight difference for τ_c, being higher than the CV max, mean, and median by .02.

B.0.2 Independent vs. ê-Conjunct

Table B.2: Cross-Validation of the Independent vs. ê-Conjunct models.

	CV Min	CV Max	CV Mean	CV Median	ME
VII					
Accuracy	67%	77%	74%	75%	73%
R2.likelihood	.08	.21	.18	.19	.17
tau.classification	.27	.49	.42	.44	.40
VAI					
Accuracy	76%	77%	77%	76%	76%
R2.likelihood	.20	.29	.27	.28	.27
tau.classification	.49	.51	.50	.50	.51
VTI					
Accuracy	70%	74%	73%	73%	72%
R2.likelihood	.16	.23	.21	.22	.20
tau.classification	.35	.45	.42	.42	.44
VTA					
Accuracy	68%	72%	71%	71%	72%
R2.likelihood	.15	.22	.20	.21	.22
tau.classification	.36	.43	.41	.42	.43

Cross-Validation was more successful in the Independent vs. ê-Conjunct alternation than in the more general Independent vs. Conjunct alternation, as seen in Table B.2. In all cases for all measures, ME models were within the CV range. All ME measures were within .02 or two percentage points of the relevant CV medians and means.

Table B.3: Cross-Validation of the Conjunct-Type models.

	CV Min	CV Max	CV Mean	CV Median	ME
VII					
Accuracy	66%	86%	77%	77%	80%
R2.likelihood	−.01	.31	.16	.17	.34
tau.classification	.24	.63	.49	.50	.56
VAI					
Accuracy	72%	76%	74%	74%	76%
R2.likelihood	.02	.11	.06	.07	.21
tau.classification	.36	.41	.39	.39	.42
VTI					
Accuracy	67%	78%	73%	73%	74%
R2.likelihood	−.02	.17	.10	.10	.25
tau.classification	.30	.45	.39	.40	.43
VTA					
Accuracy	67%	75%	71%	71%	73%
R2.likelihood	−.12	.15	.05	.07	.20
tau.classification	.32	.43	.38	.38	.42

B.0.3 Conjunct-Type

Cross-Validation showed the ME models for the Conjunct-Type alternation to be less generalizable that in other alternations, at least in terms of ρ^2 values. Although all ME models had accuracies within their respective CV range, every ME model also demonstrated ρ^2 measures that were at least .10 higher than the CV mean or medians. For the VII, VTI, and VTA classes, the CV minimums were *negative*, suggesting that the models increased badness-of-fit. For these same classes, the ME models' τ_c were always within the CV range, though the VAI ME model rested slightly outside of it (and was .03 points higher than the CV mean/median). Given the lack of a linguistic informativity offered by the Conjunct-Type alternation, the idea that this modelling is the least generalizable to new data is hardly surprising.

B.0.4 Cross-Validation Discussion

It is important to note that cross-validation with mixed-effects as in this dissertation is not strictly reliable. This is because the modelling in this dissertation uses lemma identity as a random effect and there are many lemmas in the corpus. Because the cross-validation used in this dissertation randomly samples from the corpus to create different folds, there will almost certainly be situations where a lemma is present in the general corpus, but not in a cross-validation fold.[1] When a cross-validation model then predicts a data point given a random-effect with a value/lemma identity it has not encountered, it must either report an error or compensate for this somehow. The `lme4` package's `predict.merMod` method (Bates et al., 2015) used for modelling compensates by 'use[ing] the unconditional (population-level) values for data with previously unobserved levels.' In practice, this appears to result in treatment of the missing random-effect value in a way that is unlikely to over-state or under-state its effect on modelling. As a result, if the random-effect value has a substantial impact on an outcome in the broader corpus, its modelled effect in the cross-validation fold will be non-representative of its real effect in the corpus. Put another way, the cross-validation model will not consider the impact of an unencountered random-effect value as strongly as it should if it is trying to model the general behaviour based on the overall corpus. Thus, it is possible that the large ranges for the model statistics in the Conjunct-Type alternation, along with the differences between the cross-validation models and the main mixed-effects models of this dissertation, can be explained by the nature of cross-validation when including random-effects. Finally, negative ρ^2 values in this alternation suggest that a model actually does a worse job at modelling the alternation

[1] For example, in a (separate) 10-fold random sample the Conjunct-Type subset of the corpus (not split by conjugation class), the percentage of random-effect value tokens that were missing from a fold but present in the overall Conjunct-Type subset ranged from 7% to 13% with a mean and median of 11%. For random-effect value types, these percentages ranged from 10% to 18% with a mean and median of 15%.

than it would using just verbs' relative frequencies. According to Arppe (2008, 220), this can result from cases of extremely high or low likelihoods where a small number of actually occurring outcomes are given very low probabilities.