Computational modelling of Plains Cree syntax: A Constraint Grammar approach to verbs and arguments in a Plains Cree corpus

By Katherine Schmirler

Abstract

In this paper, I introduce a Constraint Grammar parser for Plains Cree and make use of this parser to analyse the order of verbs and arguments in a corpus. Plains Cree is an indigenous language spoken primarily in the Canadian provinces of Alberta and Saskatchewan and displays non-configurational syntax. For Cree, morphological information can be used to identify relationships between verbs and arguments, as strict linear relationships cannot be determined. The rich morphology of the language generally allows for the straightforward identification of verbal arguments, though some ambiguity remains. Constraint Grammar uses morphological and syntactic information to parse sentences, making it ideal for the modelling of Cree syntax. In tandem with a morphological analyser, the Constraint Grammar parser analyses syntactic relationships within a corpus of Plains Cree and allows for the examination of phrase order on a larger scale than previously possible for most indigenous languages in Canada.

While only the orders of verbs, actors, and goals are examined herein, the results verify what has been found for Plains Cree in qualitative investigations and present opportunities for more in-depth computational research into word order in a Plains Cree corpus. Qualitative descriptions of how topic and focus may influence word order are also reflected in this corpus analysis, even though only the surface order of arguments and verb classes are represented in the results; word order and its relationships to grammatical features and information structure can also be quantitatively studied using a Constraint Grammar parser. Continued development of such computational tools also benefits language tools such as linguistically analysed texts, spell checkers, grammar checkers, and language-learning applications.[1]

1. Introduction

While the complex polysynthetic morphology of Plains Cree has been well-described for more than a century, its non-configurational syntax has proven a much more complex phenomenon to address. Various small-scale qualitative studies into the relationship between word order and information structure have been undertaken (e.g. Wolvengrey, 2011; Cook, 2014);[2] however, recent developments have made larger-scale quantitative research possible. In this paper, I discuss the development and application of a Constraint Grammar (CG) parser for the assignment of verbal argument functions in Plains Cree, using only the basic morphosyntactic features assigned to the word forms. A Constraint Grammar is used to disambiguate forms with multiple possible morphosyntactic analyses and to assign syntactic dependencies between word forms in a text. CG parsers, first designed for languages such as

[1] Thanks to Antti Arppe, Trond Trosterud, Lene Antonsen, Atticus Harrigan, Arok Wolvengrey, and Jean Okimâsis for their computational and descriptive work on Plains Cree.

[2] Such studies have examined the structure of small numbers of sentences from Cree texts, frequently referencing discourse contexts, and have elicited spoken data from consultants; they have not, however, undertaken statistical analysis beyond word order proportions.

English, use constraints that rely on relationships between words, both in terms of linear order and morphosyntactic features, to establish these dependencies.

Dependency syntax formalisms can be identified as theoretical counterparts to constraint-based parsers, which have been used to great effect to parse languages with complex morphosyntactic systems and often less restricted word orders (Sámi languages, see Trosterud 2004, 2005, 2007; Antonsen et al., 2009; Finnish, see University of Helsinki, 2017; Faroese, see Trosterud, 2009). Dependency syntax identifies surface-syntactic relationships using morphological and syntactic information. The rich polysynthetic morphology of Plains Cree combined with its non-configurational syntax make it an excellent candidate for constraint-based parsing, as linear order does not play a major role in Plains Cree, especially with respect to the order of verbs and their arguments.

Computational work on Plains Cree has developed a morphological model, using Finite State Transducer (FST) tools that can analyse and generate Plains Cree word forms (Snoek et al., 2014; Harrigan et al., forthcoming).[3] The morphological analyser outputs tag strings that represent the morphosyntactic features of the lemma and its inflections, which can then be used as input for a CG parser. In this paper, I introduce the features relevant to the recognition of verbal arguments and their dependencies in the syntax of Plains Cree and describe the basic implementation in the CG parser. The CG parser is then applied to a 72,000-word corpus of Plains Cree to determine the configurations of verbs and arguments found in the clauses therein, as a first investigation of Cree word order from a large scale computational perspective.

At this time, only overtly recognized grammatical feature tags are available for the CG to refer to; however, a Constraint Grammar, like any syntactic description, can be considerably improved by reference to lexicosemantic information and functional classification. Though beyond the scope of the current paper, supplementary lexicosemantic information is planned for future implementation in the parser.

1.1. Outline

In §2, I introduce the overt nominal and verbal features of Plains Cree relevant to syntactic roles and noun phrases, followed by descriptions of the relationships between Plains Cree verbs and arguments and within noun phrases. In §3, a brief introduction to the Dependency syntax formalism is presented and connected to the Constraint Grammar parsing formalism, the basic features of which are also described. Major features of the CG parser for Plains Cree are presented in §4, followed in §5 by the results of its application to a corpus of Plains Cree, with a focus on transitive animate verbs. Conclusions are presented in §6.

2. Plains Cree

Plains Cree is an Algonquian language spoken in Western Canada, with small numbers of speakers in northern Montana. The estimates for speaker numbers vary, from as low as 160 in the 2011 census (Statistics Canada, 2017), to 26,000 from census data in the 1970s (Wolfart, 1973) and up to 34,000 (SIL, 1982, cited in Ethnologue, 2017). For current numbers, no accurate figure is reported, though there are probably several thousands of speakers across Alberta and Saskatchewan. Books are published and radio and television programs are broadcast in Plains Cree, and children in at least some communities are learning the language; there is considerable drive among speakers to maintain and document the language. Plains Cree is the westernmost of

[3] Source files can be found at https://victorio.uit.no/langtech/trunk/langs/crk/src/morphology/.

the Cree-Montagnais-Naskapi languages, which form a dialect continuum across Canada from the Rocky Mountains to Quebec and into Labrador (Rhodes & Todd, 1981).

2.1. Grammatical features and morphology

2.1.1. Nouns and pronouns

Algonquian languages are highly polysynthetic with complex morphological systems. Plains Cree inflectional morphology is, for the most part, agglutinative with straightforward morphophonological rules. The morphosyntactic features and inflectional system of Plains Cree are modelled by means of a morphological analyser, the output of which, with morphosyntactic feature tags and all possible ambiguous analyses, becomes the input for the Constraint Grammar parser. The features discussed in this section are tagged in the output of the morphological analyser and are referenced in the constraints of the current CG parser.

Animacy is the noun classification or gender system found in Algonquian languages, with two genders, animate and inanimate. All words for people, animals, and trees are animate, while most other forms are inanimate. However, there are many words such as *asikan* 'sock' or *kôna* 'snow'[4] that are semantically inanimate but behave as animate nouns in Plains Cree. In this way, it is truly a grammatical distinction rather than a purely semantic one (e.g. Ahenakew, 1987, pp. 17-9). Sex-based gender distinctions are not made in Cree (e.g. *wiya* 'he, she, it [animate]'). Examples are given in (1) and (2).

(1) Animate nouns (Wolvengrey, 2001)

a.	*nâpêw*	'man'	*nâpêwak*	'men'
b.	*atim*	'dog'	*atimwak*	'dogs'
c.	*mistik*	'tree'	*mistikwak*	'trees'

(2) Inanimate nouns (Wolvengrey, 2001)

a.	*mâsinahikan*	'book'	*mâsinahikana*	'books'
b.	*astotin*	'hat'	*astotina*	'hats'
c.	*mîcisowinâhtik*	'table'	*mîcisowinâhtikwa*	'tables'

Nouns are inflected for number, with no marking in the singular and a suffix in the plural, -*ak* for animate and -*a* for inanimate, also given in (1) and (2) above.[5] Nouns can also be marked as locative, indicating spatial relationships such as *in*, *on*, *to*, or *at*; they often occur with prepositions or locative particles, discussed below. Locative nouns are not the arguments of verbs, nor are they marked for number or obviation (see below), though they can be possessed. There are two locative suffixes, the simple -*ihk* and the distributive -*inâhk*; the latter is used generally and the former is restricted only to nouns denoting humans and animals, and means 'among' or 'in the land of', as in (3) (Wolfart, 1973, pp. 31-2).

(3) Locatives (Wolfart, 1973, p. 31)

a.	*sâkahikan*	'lake'	*sâkahikanihk*	'at the lake'
b.	*pihko*	'ashes'	*pihkohk*	'in the ashes'

[4] Circumflexes or macrons are used to mark long vowels in written Plains Cree; all Cree forms throughout are written using the Standard Roman Orthography (SRO) (Okimâsis & Wolvengrey, 2008).

[5] While there is no synchronic marking for singular, some forms exhibit obsolete singular marking, -*a* for animate (e.g. *maskwa* 'bear') and -*i* for inanimate (e.g. *wâwi* 'egg').

c.	*paskwâwiyiniw*	‘Plains Cree’	*paskwâwiyinînâhk*	‘in Plains Cree country’
d.	*mostos*	‘buffalo’	*mostosonâhk*	‘in buffalo country’

Animate nouns are also inflected for obviation, a pragmatic category that distinguishes between more topical (proximate) and less topical (obviative) animate third persons. Plains Cree obviative third persons are marked with the suffix *-a* and do not distinguish number (Wolfart, 1973, p. 13-4). Obviation occurs, and overt marking is required, any time more than one animate third person participant is present in a discourse. This includes one animate third person acting on another, but also when one animate third person possesses another. Examples of an inanimate possessum and an animate possessum with animate possessors are given in (4). Obviation is not a case; it alone marks neither syntactic nor semantic role and an obviative noun may function as an actor or goal syntactically, an agent or patient semantically, etc., determined by verbal inflection.

(4) Possession and obviation (Okimâsis, 2004, p. 140)

a. Inanimate possessum
omôhkomân

o-	môhkomân
3	knife.IN[6]

‘his knife’

b. Animate possessum
okâwiya

o-	-kâwiy	-a
3	mother.AN	OBV

‘his mother’

Obviation can be considered part of a larger person hierarchy, which is particularly relevant in the discussion of transitive verbs below. In this hierarchy, animates are ranked over inanimates, proximate over obviative, and speech act participants (local) over non-speech act participants (non-local). Furthermore, second person is ranked over first person; this is further discussed with respect to verbs. The person hierarchy is visualized in (5).

(5) Algonquian person hierarchy (adapted from Wolvengrey, 2011, p. 57)

2 > 1 >> 3 > 3′ >> 0

Alongside nouns, pronouns also act as nominal elements as arguments of verbs. Pronouns are also inflected for animacy and number. There are personal pronouns and demonstrative pronouns; the latter are discussed in §2.2.2. The former include pronouns for animate persons: *niya* ‘1SG’, *kiya* ‘2SG’, *wiya* ‘3SG’, *niyanân* ‘1PL’, *kiyanaw* ‘21PL’, *kiyawâw* ‘2PL’, and *wiyawâw* ‘3PL’. These pronouns have various functions and can be translated in various ways; for example, *niya* can be translated as ‘I’, ‘me’, or ‘mine’. Obviative personal pronouns do not seem to occur (Wolfart, 1973, p. 38; Okimâsis 2004, pp. 19-25).

[6] Abbreviations for Plains Cree glosses: 1SG=first person singular, 1PL=first person plural exclusive, 2SG=second person singular, 21PL=first person plural inclusive, 2PL=second person plural, 3SG=third person singular, 3PL=third person plural, 3′=third person obviative, OBV=obviative, PROX=proximate, AN=animate, IN=inanimate, DIR=direct, INV=inverse, CNJ=conjunct, PST=past, FUT=future, FOC=focus, DEM=demonstrative, IPC=particle, LOC=locative, INTER=interrogative, THM=theme sign.

2.1.2. Verbs

The verbs of Plains Cree are classified by both their transitivity and the animacy of their participants. There are two classes of intransitive verbs, those with one inanimate participant, typically an actor (VII—verb inanimate intransitive) and those with one animate participant (VAI). Similarly, there are two classes of transitive verbs, though these are categorized by the animacy of their goals rather than that of their actors,[7] which are instead generally assumed to be animate.[8] Thus, VTIs (verb transitive inanimate) have inanimate goals and VTAs have animate goals.

Verbs inflect for the person, obviative status, and number of their actors, and in the case of VTAs, the person, obviative status, and number of their goals. VIIs, as they can only inflect for inanimates, inflect for third person proximate and obviative, singular and plural (obviative inanimate verbal inflections are used when animate third persons are present in the context, e.g. possessing an inanimate object which is being described). VAIs and VTIs, which both have one animate participant, the actor,[9] inflect for first, second, and third person in the singular and first person inclusive and exclusive plural, second person plural, and third person plural, as well as third person obviative.

VTAs inflect for the same categories as VAIs and VTIs, but also for the features of the second participant and the direction of the action. Algonquian languages have direct-inverse systems, where rather than coding for subject or object (e.g. as done in case-marking languages), the direction of an action is marked. The person and number marking does not change depending on the role of each participant; the morphology only indicates the person and number features of the participants, not their roles. For example, in (6) below, the marking for second person and third person plural do not change, but the direction theme sign indicates which is the actor and which is the goal (Wolvengrey, 2011, pp. 173-6).

(6) Direction marking (Wolvengrey, 2011, p. 174)

a. *kiwîcihâwak*

ki-	wîcih-	-â	-wak
2	help.VTA	DIR	3PL

‘You help them.’

---

[7] *Actor* and *goal* are terms frequently used in Algonquian literature; the status of syntactic roles such as subject and object is uncertain in Cree morphosyntax. I follow the Algonquianist convention throughout (e.g. Bloomfield, 1946; Wolvengrey, 2011). Together, these may be the called arguments or participants; I use both terms interchangeably in this paper.

[8] While this criterion is sufficient to describe most uses of VTAs, there is also a minor inflectional category of inanimate actor using VTA stems, indicating when an inanimate participant acts on an animate one (Wolfart, 1973, p. 61-2). These inflections are not yet implemented in the morphological model but have been included in the manually-verified corpus.

[9] Wolvengrey (2011) proposes an alternative verbal classification system with this very issue at its core. Overall, there are very few differences between VAI and VTI inflections, while VII and VTA are as different from VAI/VTI as each other. The system proposed by Wolvengrey (2011) consists of three classes: V0, for VIIs with zero animate participants, V1, for VAIs and VTIs with one animate participant, and V2, for VTAs with two animate participants. For the purposes of the morphological analyser, we have maintained the traditional four-part classification system, which can carry key inflectional and semantic information regarding the number of possible arguments, which feeds into the CG.

b. *kiwîcihikwak*

ki-	wîcih-	-ikw	-wak
2	help.VTA	INV	3PL

‘They help you.’

The choice of direct or inverse morphology is determined on the basis of the person hierarchy, given above. When a higher-ranked participant acts on a lower-ranked participant, the morphology is direct. In the opposite case, the morphology is inverse.[10]

As mentioned above, the classification of obviation as pragmatic rather than syntactic role marking becomes apparent with respect to direction marking in the VTAs. In (7) below, there is no change in the obviative marking on the nouns, but only in the direction marking on the verb; in doing so, the semantic roles are switched. Similarly, if one were to change only which noun is obviative and leave the direction unchanged, the roles would again be reversed. Grammatical features such as person, number, obviation, and direction are key to finding links between verbs and their arguments words in Cree sentences.

(7) Plains Cree (adapted from Wolvengrey, 2011, p. 175)

a. *câniy kî-wîcihêw mêrîwa*

câniy	kî-	wîcih-	**-ê**	-w	mêriy	-wa
Johnny.PROX	PST	help.VTA	DIR	3SG	Mary	OBV

‘Johnny helped Mary.’

b. *câniy kî-wîcihik mêrîwa*

câniy	kî-	wîcih-	**-ik(w)**	-(w)	mêriy	-wa
Johnny.PROX	PST	help.VTA	INV	3SG	Mary	OBV

‘Mary helped Johnny.’ / ‘Johnny was helped by Mary.’

Verbs in Plains Cree occur in one of several verbal orders, or patterns of inflection, each with different morphology and function. The verbs exemplified above are in the independent order, which is used for matrix clauses (and, crucially, cannot occur in embedded clauses). They are used especially when there is no previously established (linguistic or extralinguistic) context, and are often citation forms. The other two orders are the conjunct and the imperative. Imperative verbs can be either immediate or delayed and inflect for second person singular, second person plural, and the first person plural exclusive. Conjunct verbs have two subclasses, simple and subjunctive; the latter is also known as the future conditional.[11] In modern Plains Cree, the conjunct order frequently contains a prefix unique to conjunct verbs and all person marking occurs via suffixation. Older forms of Plains Cree made more use of “changed conjunct” forms, wherein alterations were made to the first vowel; this has been supplanted by

---

[10] Many a linguist has been tempted to describe the direct-inverse distinction in terms of a perhaps more familiar active-passive one. However, it is important to note that the inverse is the *only* means of indicating a first person acting on a second person; this is an active form, not a passive transformation of an underlying structure. Furthermore, using tests for objecthood and subjecthood, Dahlstrom (1991) demonstrated that inverse verbs are active, not passive; the only difference is that a less topical participant is the actor. Still, when translated into English, passives are often used to indicate the topical status of the goal/patient, especially in situations where both participants are lexicalised.

[11] The morphological analyser for Plains Cree contains tags for conjunct, meaning only the simple conjunct, and future conditional.

the neutral conjunct preverb *ê-*, though some forms do still occur in the corpus. The future conditional is often translated as 'if or when…' and is formed using the conjunct suffixes with a few minor differences followed by an *-i* or *-o* suffix (Wolfart, 1973). Each of these orders is included in the morphological model's morphosyntactic tag output; however, as interclausal relationships are not implemented in the CG at this point, they are not discussed further. More information can be found in Appendix A.

Alongside nominals and verbs, there is one other class of words in Plains Cree. These are particles; this heterogeneous class contains all other words in Plains Cree, which, unlike nouns and verbs, are indeclinable. Subclasses, though not yet formally classified, can be identified: for example, one can readily identify interrogative markers, prepositions, temporal adverbials, spatial adverbials, quantifiers, numerals, and more.[12] There are some particles that are prenominal or preverbal (prenouns or preverbs), that occur prefixed to nouns or verbs and modify their meanings; these may be grammatical (e.g. tense) or lexical, expressing adverbial meanings. Some lexical preverbs may also be used as free particles (Bloomfield, 1946; Wolfart 1973; 1996; Wolvengrey, 2001). There are also particle phrases or phrasal particles, wherein a string of two or more particles constituent one semantic unit. These are included in Wolvengrey (2001), but are not yet implemented in the Plains Cree morphological analyser.

A key point that arises in the available syntactic descriptions is the importance of lexical semantics in Plains Cree syntactic relationships: the morphosyntactic feature tags output by the morphological model will only allow the syntax to be modelled to a certain degree.[13] Though this paper only introduces a fledgling CG for Plains Cree, in order to model the syntax thoroughly, at least some reference to lexical semantics must be made in the future. Various such relationships are discussed in Appendix A; some comments also appear in the discussion of the results.

2.2. Syntactic descriptions

2.2.1. Verbs and arguments

While Cree morphology is well-described, much less is established regarding the syntax of the language. The complex feature-marking on verbs as described above allows Cree to be a non-configurational language,[14] in the sense that any order of the subject, verb, and object is semantically (if not pragmatically) equivalent to any other order, as in (8). Word order has no bearing on the semantic roles of 'children' and 'ducks'; these are achieved through obviation on the nouns and direction morphology on the verb.

(8) Plains Cree: 'the children killed some ducks' (adapted from Wolfart, 1996, p. 392)

SVO:	*awâsisak nipahêwak sîsîpa*	(children killed ducks)
SOV:	*awâsisak sîsîpa nipahêwak*	(children ducks killed)
VSO:	*nipahêwak awâsisak sîsîpa*	(killed children ducks)

---

12 Oxford (2007) classified particles of Innu Aimun, an Eastern Cree dialect, in terms of their functions. The categories include adnominal (adjectives, quantifiers), prepositions (locative, non-locative), adverbs (semantic functions, syntactic positions—e.g. phrase vs. sentence-level), focus particles, question particles, negators, conjunctions (coordinating, subordinating, conjunctive—i.e. discourse-level connections), and interjections. These categories would offer an excellent template to begin similar classification of Plains Cree particles for inclusion in a parser or morphological analyser, though such classification is beyond the scope of the present paper.

13 Various investigations of Plains Cree syntax make considerable reference to lexical semantics (Dahlstrom, 1991; Hirose, 2003; Cook, 2014; Wolvengrey, 2011; Blain, 1997, to name but a few).

14 Free word order and flexible word order are also labels that can be used to describe this phenomenon in Plains Cree. Non-configurational is used throughout.

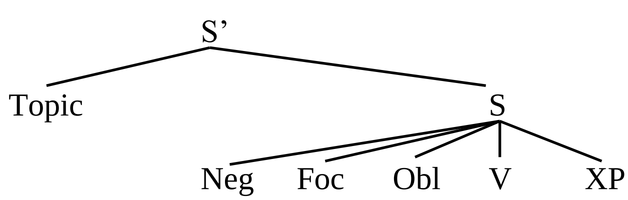
VOS:	*nipahêwak sîsîpa awâsisak*	(killed ducks children)
OVS:	*sîsîpa nipahêwak awâsisak*	(ducks killed children)
OSV:	*sîsîpa awâsisak nipahêwak*	(ducks children killed)

While each of the above orders is possible, the extensive verbal morphology does not require overt arguments, there are also transitive clauses that contain SV or VS, VO or OV orders, or simply the verb without any lexicalised arguments (Wolfart, 1996); lexicalised arguments, i.e. overt nouns and pronouns, have also been described as adjuncts to the verb within generative formalisms (e.g. Blain, 1997).[15] Wolvengrey (2011) investigated the combinations of subjects, verbs, and objects for monotransitive VTAs as found in one collection of Cree texts, *Stories of the House People* (Vandall & Douquette, 1987), abbreviated hereafter as the HP texts, and reported frequency of each word order permutation. These included clauses where one or both arguments were lexicalised, as well as the clauses with no overtly lexicalised arguments, for a total of 286 clauses.

In HP texts, the six possible word order permutations where both participants are lexicalised make up only a small percentage of the total clauses, ranging from 0.7% for VOS (n = 2) to 3.8% for OVS (n = 11). Clauses with only a lexicalised subject are more common: SV occurs in 3.8% of cases (n = 11) and VS in 5.2% (n = 15). It is even more common for only the object to be lexicalised: OV occurs in 18.9% of cases (n = 54) in VO in 29.7% (n = 85). Clauses in which neither participant is lexicalised (i.e. there is only the verb) occurred in 30.8% of cases (n = 88). As these numbers describe only one Cree text, they may not reflect counts that might be found in a larger corpus (Wolvengrey, 2011, p. 202). Mühlbauer (2007) suggests that VSO is the basic or neutral word order in Cree; while this order, with both arguments lexicalised, occurs in only 1.0% of the clauses examined in Wolvengrey (2011), VS and VO are considerably more common (see numbers above).

Though the various possible word orders are considered semantically equivalent, they can still be influenced by pragmatic factors. Across Algonquian languages, the first position of a clause frequently contains a topical argument, followed by negation, focus, oblique elements, and then the verb, as in the template given in (9). The post-verbal material in a clause, however, does not present agreed-upon patterns.

(9) The Algonquian clause (Dahlstrom, 1995, p. 3)

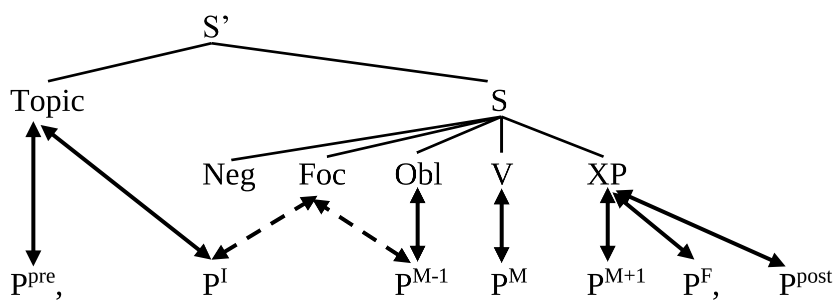


Wolvengrey (2011) expands upon this template using the Functional Discourse Grammar (FDG) framework, which orders elements with respect to a clause-medial position (P^{M}), which for Plains Cree is occupied by the verb. Other fixed positions include P^{pre}, a pre-clausal position, which Wolvengrey (2011) associates with the topic in the above tree diagram, P^{I}, the initial position in the clause, which is associated with both topic and focus, and P^{M-1}, which

[15] Personal and demonstrative pronouns are also not required with verbs, and when they do occur as arguments, they serve emphatic functions.

immediately precedes P^{M} and contains either a focus element or an oblique. Further positions, associated with the post-verbal XP above, include P^{M+1} (immediately following P^{M}), P^{F} (final element in the clause), and P^{post} (post-clausal elements). These correspond to the above tree diagram as given in (10).[16] A Plains Cree sentence exemplifying such a template is given in (11). The outer clause containing *awînihi anihi* indicates a cleft structure.

(10) Clausal positions in FDG (Wolvengrey, 2011, p. 211)



(11) A Plains Cree sentence in an FDG clausal template (adapted from Wolvengrey, 2011, p. 335)

P^{I}	P^{2}	[		P^{M}		]
awînihi	*anihi*	[P^{I}	P^{I+1}	P^{M-1}	P^{M}	]
		[*otâkosîhk*	*câniy*	*mînisihk*	*kâ-kî-wâpamikot*	]
who.OBV	that	yesterday	John.PROX	Saskatoon	CNJ-PST-saw.3′>3SG	

'Who was it that saw John in Saskatoon yesterday?'

Wolvengrey (2011) also explores the post-verbal portion of the clause, placing arguments in post-medial positions (e.g. P^{M+1}, P^{M+2}, etc.) in verb-initial clauses, which are considered an unmarked order for many Algonquian languages. For transitive verbs, the highest-ranking argument (in terms of semantic role, e.g. AGENT>RECIPIENT>PATIENT, etc.) is more likely to be first, though this is only a tendency. Ditransitive VTAs display a more consistent pattern; when both the goal/recipient and indirect object/oblique are animate, the grammatical information (animacy, obviation) alone is not sufficient to determine their respective roles. Wolvengrey (2011) presents the following examples and discusses comments made by Cree consultants.

(12) Ditransitive post-verbal restrictions (Wolvengrey, 2011, pp. 213-4)

a. *ana nâpêw kî-asamêw atimwa kinosêwa*.

ana	nâpêw	kî-	asam	-ê	-w	atimw-
DEM.3SG	man.NA	PST	feed.VTA	DIR	3SG	dog.NA
-a	kinosêw	-a				
OBV	fish.NA	OBV				

'That man fed (a/the) dog(s) fish.'/'That man fed fish to (a/the) dog(s).'

[16] Solid lines indicate that a certain functional role will displace one connected by a dotted line. Hence, when a Topic is displaced from a pre-clausal position, it will take precedence over focus in P^{I}, displacing focus (if it occurs) to a later position.

b. *ana nâpêw kî-asamêw kinosêwa atimwa*.

ana	nâpêw	kî-	asam	-ê	-w	kinosêw
DEM.3SG	man.NA	PST	feed.VTA	DIR	3SG	fish.NA

-a	atimw-	-a
OBV	dog.NA	OBV

'That man fed (a/the) fish dog.'/'That man fed dog to (a/the) fish.'

In the examples presented in (12), we see that word order in Plains Cree can be used to determine semantic roles in cases where direction and obviation are insufficient; the recipient or direct object occurs before the patient/theme or direct object.[17] However, pragmatic information can overrule such syntactic information. For example, in (13), a possessed form of 'dog' is used,[18] indicating a closer relationship between the actor and another entity in the clause. When this occurs, ranking of semantic roles with respect to word order is demoted in favour of pragmatics; speakers interpret both word orders the same way because it is pragmatically unlikely for a person to feed their own pet dog(s) to a fish. While (13)b. can be interpreted as a pet dog being fed to a fish, speakers find this unlikely (and rather humorous) and would avoid the word order.

(13) Pragmatic post-verbal interpretations (Wolvengrey, 2011, p. 215)

a. *ana nâpêw kî-asamêw otêma kinosêwa*.

ana	nâpêw	kî-asam	-ê	-w	o-	-têm-	-a
DEM.3SG	man.NA	PST-feed.VTA	DIR	3SG	3SG	dog.NA	OBV

kinosêw	-a
FISH	OBV

'That man fed his dog(s) fish.'/'That man fed fish to his dog(s).'

b. *?ana nâpêw kî-asamêw kinosêwa otêma*.

ana	nâpêw	kî-asam	-ê	-w	kinosêw	-a
DEM.3SG	man.NA	PST-feed.VTA	DIR	3SG	fish.NA	OBV

o-	-têm-	-a
3SG	dog.NA	OBV

'That man fed his dog(s) fish./That man fed fish to his dog(s).'

As these examples show, word order can have some bearing on determining syntactic and semantic roles in Plains Cree. However, this rule cannot be categorically applied to Plains Cree sentences for two key pragmatic reasons. First, we see that pragmatics can overrule the syntax to allow for the most likely interpretation. Second, as many Cree arguments are not lexicalised, instances where both the recipient/goal and indirect object are present will be rather rare. While word order is not as important a factor in constructing the CG, arguments can be determined by grammatical marking such as direction and obviation.

---

[17] Interestingly, this pattern is the opposite of that seen in Innu, an eastern dialect in the Cree-Montagnais-Naskapi continuum, where in ditransitive constructions the indirect object occurs before the direct object (Branigan & MacKenzie, 1999).

[18] When 'dog' is possessed, rather than using the free stem *atimw-*, the bound form *-têm-* which can mean 'dog' or 'horse' is used instead.

2.2.2. The noun phrase: nouns, pronouns, and demonstratives

Though the order of the verb and argument(s) in Plains Cree clauses is not fixed, stricter orders can be determined for other syntactic relationships. These syntactic relationships are important not only for determining semantic relationships in the CG, but also for disambiguating forms with multiple analyses; such ambiguity exists between demonstrative pronouns and determiners.[19] For example, the demonstrative *ôma* can mean ‘this (one) [inanimate]’ with no overt noun or it can occur with an overt inanimate noun. When it occurs with a noun, its position relative to that noun determines the meaning of the phrase: *ôma masinahikan* ‘this book’, but *masinahikan ôma* ‘this is a book’. Additionally, a subset of demonstratives can also serve as focus markers, indicating that a preceding noun or pronoun is focused. This allows constructions such as the following to occur, wherein *awa*, the animate counterpart to *ôma*, can have various functions (14).

(14) Demonstratives and focus (Wolvengrey, 2011, p. 296)

aw âwa mahihkan

i. awa awa mahihkan
this.AN this.AN wolf.AN
‘This is the wolf.’

OR

ii. awa awa mahihkan
this.AN FOC wolf.AN
‘this here wolf’

Wolvengrey constructs extreme examples wherein forms such as *ôma* or *awa* can be used in a single sentence with all three functions, such as *ôm ôm ôma* ‘this is the one here (inanimate), where the functions of *ôma* are demonstrative pronoun, focus marker, and predicating demonstrative respectively (2011, pp. 296-7).

Similar to the predicating function of post-nominal demonstratives, personal pronouns can also predicate nouns in this manner. As with demonstratives, no copula is used for predication. However, a verbal construction can be used instead; a VAI stem is derived via the suffix -*iwi* and inflected as usual. These constructions are exemplified in (15).

(15) Copular constructions (Wolvengrey, 2011, pp. 294-5)

a. *mahihkan niya*.
mahihkan niya
wolf.AN 1SG
‘I am Wolf (a name).’/‘I am a wolf.’

---

[19] Demonstratives occur in proximal, medial, and distal forms, singular and plural, animate and inanimate, and proximate and obviative for the animate forms. The proximal inanimate singular is *ôma*; *awa* is the animate counterpart. Animate obviative pronouns are identical in form to plural inanimate pronouns, introducing ambiguity. The syncretism between the animate obviative and the inanimate plural can also be seen in the nominal suffix -*a*; the syncretism in the pronouns appears to be found in Proto-Algonquian (Proulx, 1988).

b. *nimahihkaniwin*.

ni-	mahkihkaniwi-	-n
1	be.a.wolf.VAI	1/2SG

'I am a wolf.'

Personal pronouns and demonstratives may also occur together in such phrases, with differing meanings dependent upon the animacy of the demonstrative, as in demonstrated in (16). In (16)a., the pronoun indicates possession of the demonstrative; in (16)b., the same is true, but there is also a nominal antecedent for the third person pronoun, *wiya*. For (16)c., the form *awa* focuses 'wolf', which is possessed by the first person pronoun, the position of which predicates 'wolf'.

(16) Demonstratives and pronouns (Okimâsis, 2004, p. 20; Wolvengrey, 2011, p. 295)

a. *niya anima*.

niya	anima
1SG	that.IN

'That (inanimate thing) is mine.'

b. *mahihkan ôma wiya*.

mahihkan	ôma	wiya
wolf.AN	this.IN	3SG

'This (inanimate thing) is Wolf's.'

c. *mahihkan awa niya*.

mahihkan	awa	niya
wolf.AN	FOC	1SG

'This wolf (here) is mine.'

The syntactic relationships between nouns, personal pronouns, and demonstrative pronouns allow us to begin disambiguating between the various functions of demonstrative pronouns and determine the semantic relationships within phrases involving nominal and pronominal elements. These noun phrases can then be related to verbs as their arguments in the form of dependencies.

There are similar ordering principles concerning possession. In cases of possession, the possessor is the more topical or focused element (and, in the case of two animate nouns, the possessor is the only one that can occur as proximate), and as per the clausal templates laid out above, the possessor is then more likely to occur first (Wolvengrey, 2011, p. 16; Ahenakew, 1987, p. 33). The overt third person possessor also acts as an antecedent to the possessive morphology (e.g. Cook, 2014), but if the possessor is already topical, it does not need to be specified in the phrase itself. Features of several elements in phrases involving possession can be used to determine semantic relationships: firstly, the morphosyntactic features of the possessed noun are likely to indicate that it is an argument of the verb (and the features of the possessor will not match, e.g. in animacy, number, obviation), secondly, the number of an unmarked noun, and thirdly, the order of an unmarked noun and a possessed noun, may be used to determine possessive relationships. Very few instances of these relationships are found in the corpus and are not yet implemented in the parser, and so they are not commented on further here.

3. Dependencies and constraints

3.1. Features of dependency syntax

Dependency syntax is a formalism for codifying relationships between elements in a sentence with respect to syntactic, semantic, and morphological relationships between words, rather than the immediate constituency of phrase structure grammars. Dependency analyses focus on relationships between words at all levels, not their linear order as a transformation of a presumed underlying representation. Rather than belonging to trees or phrases, individual words are related to other forms in a sentence. In dependency tree representations, every node is a terminal node; relationships between nodes are dependencies, not dominance. Unlike phrase structure representations, dependency grammar does not attempt to represent or determine linear order of word forms, making it ideal for the computational analysis of non-configurational languages such as Cree (Mel'cuk, 1988).

Information can be coded in language in various ways. Lexemes carry meaning, and the order in which they occur, the intonation patterns they occur with, and the inflectional marking that is present may all express some portion of the overall message. These means of expression interact with semantics and syntax: lexemes can have semantic content or be functional, word order can express semantic roles or syntactic relationships, intonation can mark illocutionary force or phrase boundaries, and inflections can express meaning or occur only to mark agreement (Mel'cuk, 1988, p. 19-20).[20] Word order is a major category for expressing relationships; even in Cree, a non-configurational language, examples in the previous section have demonstrated that the order of word forms has considerable influence on syntactic and semantic relationships at least within noun phrases, even if order alone cannot be used to determine the semantic roles of verbal arguments. Similarly, it is unsurprising that semantics can play a considerable role in dependencies, given the number of issues discussed in Appendix A that can be solved with semantic information.

In dependency syntax, binary relationships between word forms are syntactic dependencies, in which one member is the governor or head and the other is the dependent. Mel'cuk (1988, pp. 21-2) lays out several required characteristics of syntactic dependencies. One, they must be *antisymmetric*: two word forms cannot govern/depend on each other. Two, they must be *antireflexive*: a word form cannot govern itself. Three, they must be *antitransitive*: the governor of a word form cannot also govern the dependent of that word form (if X→Y and Y→Z, then X cannot govern also Z). Four, each syntactic relation is *labelled*: it serves a particular function that can be distinguished from that of other syntactic dependencies. Note that these criteria describe only syntactic dependencies; morphological and semantic dependencies also occur, and these account for apparent contradictions to the above, such as agreement marking.

Though dependency trees do not code linear order as such, facts about word order can often be gleaned and may be relevant to identifying syntactic dependencies. Mel'cuk (1988, pp. 35-8) discusses the notion of *projectivity*: most sentences (at least in the languages he cites) are projective, which means that dependency arcs do not cross each other or cross over the top node (the head of a phrase, such as the verb in a clause). When dependencies cross each other, this

[20] Dependency syntax, while generally sufficient for most syntactic relationships, must also include subtrees or groupings that function as wholes. Mel'cuk (1988, 31-3) concedes that these phrase structure-like elements must occur for the sake of some constructions, especially coordinated groupings. However, he stresses that these are not constituents, as they still have full dependencies within them and they are not dominated by a node as in phrase structure approaches.

does not entail ungrammaticality, but can instead be used for stylistic marking or emphasis, such topic and focus marking as described for Plains Cree above. Furthermore, in the Plains Cree corpus, examples are found where demonstratives and nouns occur on either side of the verb, but share agreement features and so both may be recognised as the same argument of the verb. Under the current analysis, both are dependent on the verb, though it may not be unreasonable to instead recognise the relationship between the demonstrative and the noun. The current analysis may, however, better represent topic or focus in Plains Cree.

Underlying the syntactic, morphological, and semantic dependencies of dependency syntax is the Meaning-Text Model (MTM). In its most basic form, this model is concerned with how semantic content (meaning) is mapped to linguistic material (text) and vice versa (Mel'cuk, 1988, pp. 42-7). It can be expanded to include several levels of representation: semantic, deep syntactic, surface syntactic, deep morphological, surface morphological, deep phonetic, and surface phonetic, where deep refers to meaning and surface to what is realised from that meaning. These seven levels of representation can be simultaneously applied to any utterance and interact with each other to express semantic content (Mel'cuk, 1988, pp. 48-51). As linear order need not be realised in either the semantic or deep syntactic representations, this model as the underpinning of dependency syntax is particularly well-suited to the analysis of non-configurational languages, especially as a very similar pattern occurs in Functional Discourse Grammar (e.g. Hengeveld & Mackenzie, 2008), which is the formalism explored in Wolvengrey (2011) for Plains Cree. In Functional Discourse Grammar, there is a similar top-down hierarchy of levels, the Interpersonal (pragmatic) Level, the Representational (semantic) Level, the Morphosyntactic Level (clausal, phrasal, and morphological templates), and the Phonological Level. While these do not correspond to the levels of representation in the Meaning-to-Text model, the similarities between the two formalisms are unsurprising given their historical development: both have roots in the Prague School of Syntax (Graffi, 2001, pp. 195-7, 311-2), itself influenced by the work of Tesnière (1959), which focused on hierarchical connections between syntactic elements. While the above-cited works contain detailed description and exemplification of semantic, syntactic, and morphological representations, I do not go into further detail of these formalisms herein.

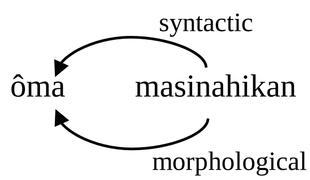
For the purposes of CG, the surface syntactic and deep morphological representations, and syntactic, morphological, and to some extent semantic, dependencies invite further discussion. The surface syntactic and morphological representations serve as the input to the CG: the surface order of sentences in the text is analogous to the surface syntactic representation, and the morphological tags from the analyser to the deep morphological representation. It is at these levels that syntactic and morphological dependencies become apparent. Morphological dependencies occur with respect to morphological features (not their exponence); a word is morphologically dependent on another when its features are determined by its governor. As some words are invariable in terms of their features, they cannot be morphologically dependent on another element (e.g. function words), though ideally all words would be involved in some syntactic dependency in a full dependency analysis.[21] Unlike syntactic (and semantic) dependencies, morphological dependencies can be symmetric, wherein two words depend on each other for their morphological features, e.g. agreement on a verb and case on a noun (Mel'cuk, 1988, pp. 107-16).

---

[21] Some words, such as *perhaps*, *maybe*, *unfortunately*, may operate at various levels, modifying words, clauses, sentences, or larger sections of discourse, and as such determining dependencies may not be straightforward.

Syntactic dependencies are built upon the finite lists of the roles that certain forms are able to take in a sentence, or their surface-syntactic *passive valency* (the roles they can take as governed by other elements, not the roles they govern, such as the valency of a verb). For example, nouns can be subjects, objects, prepositional complements, etc. When one word is syntactically dependent on another, they form a unit (a phrase) with the passive valency of the governor or head (Mel'cuk, 1988, pp. 112-3). A Plains Cree example is given in (17) below; here, *masinahikan* ‘book’ both syntactically and morphologically governs (note the direction of the arrows) *ôma*, its modifier, as the phrase can behave as a noun syntactically, e.g. be the argument of a verb (as an NP) and the form of the demonstrative is governed by the animacy and number of the noun (inanimate, singular).[22]

(17) Plains Cree demonstrative + noun



‘this book’

Syntactic dependencies may be determined by linear or prosodic correlation between elements, the determination of passive valency, and morphological links. A prosodic correlation can involve any prosodic unit, which is intentionally left vague for the purposes of the formalism; two word forms that can form a prosodic unit are linked by a syntactic dependency. For linear correlation, two word forms are linked via syntactic dependency if the position of one is determined with respect to the position of another. For example, a preposition that occurs before a noun; in this syntactic dependency, the preposition governs the noun, determines the overall passive valency of the phrase (e.g. the pair behaves like a preposition, not a noun, i.e. a prepositional phrase), and, in the case of Cree, would also entail a morphological dependency of the noun on the preposition, as the locative suffix is required on the noun. Even when there is no linear correlation, the imposition of passive valency will also involve a syntactic dependency: whenever one element determines the role (passive valency) of another, a syntactic dependency occurs. For example, regardless of word order, arguments are syntactically dependent on their verb, and as a group these can then form a larger unit that can be coordinated, subordinated, etc. (e.g. a verb phrase). Morphological links can be used to determine the head of a phrase; the head or governor within a phrase linked by a syntactic dependency could still be morphologically dependent on an element outside of that phrase (Mel'cuk, 1988, pp. 129-40). These units, regardless of the contiguity of their members, can and be referred to as phrases (e.g. noun phrase, prepositional phrase, verb phrase.

Though this background on dependency syntax is brief and touches only upon those details necessary for straightforward modelling of Plains Cree syntax, the formalism is much more complex than has been presented here. In the following subsection, I discuss Constraint Grammar with reference to the features of dependency syntax I have laid out above.

3.2. Constraint Grammar

A Constraint Grammar (or Constraint Grammar Parser) is a list of descriptive, context-based constraints designed to parse natural language. The input to the list of constraints is a morphologically-tagged corpus; in the case of Plains Cree, this corpus has been analysed by the

[22] In dependency syntax, arcs or arrows with labels are used to mark dependencies.

morphological analyser and manually verified to create a morphological "Gold Standard" corpus. The constraints disambiguate forms on the basis of morphological and lexical information to output a single surface morphosyntactic reading for each utterance (Karlsson, 1990). Unlike a generative approach to grammar, no input is considered incorrect or ungrammatical; every input is instead analysed as best as the parser allows, regardless of its grammaticality. Like dependency syntax, it looks at the surface syntactic structure and does not refer to constituents or underlying syntactic representation. Syntactic dependencies involve heads and modifiers based on morphological and linear factors; labels for these are discussed below. The fundamental underpinnings of Constraint Grammar are the rules of a language that a linguist can easily identify as categorical, and do not need to be learned as tendencies; the constraints can then be tested against a corpus of natural language to determine how accurate the rules identified by linguists are (F. Karlsson, personal communication to A. Arppe).

The first step to create the input for the CG is morphological analysis, such as that returned by a morphological analyser. The morphological analyses offered for each word form are the *readings* in its *cohort*. The morphological analysis stage is preprocessed to optimise the application of the CG; ideally, idiomatic phrases are marked as single units, and some ambiguities can be dealt with by choosing the reading with the least morphological complexity, or by combining readings where the combination does not obscure other information (i.e., underspecify the form).[23] With the morphological analysis and preprocessing complete, the CG has two main goals: 1) the disambiguation of remaining forms with a cohort of more than one morphological analysis and 2) the assignment of syntactic functions (e.g. dependencies) using the sentential context of each word form. CG can delineate the range in which to look for dependencies by referencing clause boundary punctuation in the text, such as periods, question marks, exclamation points, commas, semicolons, quotation marks, etc. Constraints are used to narrow down (disambiguate) the relationships between words in a sentence to return an analysis (Karlsson, 1995a).

The constraints in CG are of several different types. There are constraints that disambiguate based on the context, constraints that map the clause boundaries using punctuation and capitalization, and constraints that map the syntactic functions of word forms (like dependencies).[24] While the constraints can be divided into different types, they can be structured so that they occur simultaneously and interact in the parser, or structured so that groups of constraints apply sequentially. Karlsson (1995b) lists several steps to the CG. First, clause boundary information can be introduced in the input, especially within sentences, by both punctuation/capitalization and through lexical items that are known to introduce clauses, such as English *if*, *when*, *which*, etc. The input to the CG can already contain syntactic labels, if the function of a word form is completely unambiguous. This information must already be included in the lexicon, but simplifies the constraints needed. For example, the English article *a* can be marked as DN>.[25] This is followed by the disambiguation of word forms based on the context, then clause boundaries may be identified again because of this disambiguation. This cycle may be repeated, or the repeated identification of clause boundaries may be omitted entirely.

---

[23] Karlsson (1995a) uses an example in Swedish, where verbs may agree with either definite or indefinite nouns. However, these have the same verbal morphology and so these are combined to be underspecified for nominal definiteness to simplify the cohort that is examined by the CG.

[24] Constraint Grammar is well-designed for the creation of grammar checkers, and so is oriented to written, well-punctuated language (e.g. Arppe, 2000; Birn, 2000).

[25] A D(eterminer) dependent on a N(oun) that is to the right (>). Note that this particular labelling system is not carried into the Plains Cree CG parser.

Additionally, alongside the automatic disambiguation, the CG may require manual disambiguation of word forms that the grammar cannot disambiguate on basis of context. Once as many clause boundaries as possible are identified, and the ambiguities dealt with, then the morphological information and word order information are mapped to syntactic roles, resulting in a complete surface morphosyntactic analysis.

Alongside context-based constraints, heuristic-based parsing can also be used to improve the analysis. Heuristic constraints may be used to identify morphological features, or to disambiguate where normal constraints are insufficient; they can be used to reapply the same disambiguation throughout the text when context is insufficient, and to simplify issues by ignoring constraints and enforcing analyses. They can also be purely probabilistic and choose the analyses that is the most likely based on quantitative analysis. Such heuristic constraints can be used at all levels, interleaved with the usual context-based constraints (Karlsson, 1995b). Though not yet widely implemented for Plains Cree, heuristic constraints would prove useful for a number of frequent ambiguous forms for which context is not sufficient for disambiguation (see below).

The constraints themselves are written in a file of up to thirteen different sections. A number of these define sets that the constraints refer to: sentence delimiters, grammatical properties such as the word (sub)classes and grammatical features the constraints reference, and the syntactic functions that will be applied. Templates can be specified, which designate variables that can be used to simplify constraints later in the file. Morphosyntactic mappings assign syntactic function labels based on morphological features by means of lists with the following format given in (18)a:

(18) Syntactic function label constraint

a. <target, context condition(s), syntactic label(s)>
b. <Pron ((0 NOM)(1 VFIN)) @SUBJ>

The target will include the morphological features of the word (such as a part of speech or inflectional feature) the label will be applied to, the context conditions are its position relative to another form with designated features, and the syntactic labels are those that will be applied. In the simple example in (18)b., the pronominal subject of an English verb can be labelled using such a constraint: target a pronoun, such that the context conditions include that the pronoun itself (0) is in nominative case and immediately to the right (1) there is a finite verb, and then assign the syntactic label @SUBJ for subject. The format used in the current CG places the label at the beginning, but is otherwise the same.

Morphosyntactic mappings are followed by the disambiguation constraints, which have four obligatory constituents and one optional constituent. The domain, target, operator, and context condition(s) are obligatory, while the indication of a clause boundary is optional. The domain may be a particular word form or a variable representing the features of the target, which is the reading (feature or group of features) the constraint is about. The operator indicates what the CG parser is supposed to do when it finds a target with the designated features in the stated domain: if the conditions are met, either the target can be chosen and all other readings discarded, or the target itself can be discarded, and if the conditions are not met, the target can be discarded. The context is indicated using a list such as that in (19):

(19) Context condition template
a. <polarity, position, set>

The polarity is either negative (NOT) or positive (nothing). The position is defined with respect to the target, which is considered position 0. The position of a word form in the context is indicated using numbers: 1 is one word to the right, -2 is two words to the left, etc. It is not necessary to indicate absolute positions; an asterisk may be used to indicate that the word may be within any range to the left or right of the target, which is a useful feature for parsing non-configurational languages. The set is one of the set names that was declared above; the word form in the designated position is expected to have the feature(s) of that set in one reading of its cohort. Compare to (18)b. above, the context conditions ((0 NOM)(1 VFIN)) indicate that the target pronoun itself must be nominative, and that one position to the right there is a finite verb. Thus, (1 N) indicates that the operator should be applied if there is a noun immediately to the right of the target and (NOT *-1 V) indicates that the operator should be applied if there is not a verb anywhere to the left of the target (within the clause boundaries). Several conditions can be used for each constraint, and if different conditions refer to the same word form (e.g. (1 N)(1 NOM)—a nominative noun), all conditions must be met to accept or discard the target reading. However, a form does not need to be explicitly selected to be the preferred form; furthermore, if a constraint would rule out the las remaining or only member of a cohort, the constraint does not apply (Karlsson, 1995b).

Much like the morphological Gold Standard corpus that has been manually verified for Plains Cree, the results of CG parsing can also be manually verified: this improves the analysis, can be used to identify areas for improvement in the parser, and can be used to train the parser for future disambiguation. As stated above, the goal of this paper is to describe the set of Plains Cree syntactic rules that can be implemented on the basis of morphosyntactic features only, creating a far from exhaustive parser that will form only the basis for a comprehensive parser for Plains Cree. As such, only a smaller portion of the text has been manually coded; this is described below.

4. A Constraint Grammar for Plains Cree

As mentioned above, the input to the Constraint Grammar is similar to the surface syntactic representation of dependency syntax: the text contains words in their surface order, and the manually-verified morphological Gold Standard analysis contains word class information (e.g. valency of verbs, animacy) and morphological feature tags.[26] One can then make use of features and word order to establish links (dependencies) between word forms using constraints that contain relative position and required tag information. Clause boundaries are used to delimit the range over which a constraint can operate; sentence-level punctuation such as periods, question marks, and exclamation points are included, as well as commas.

Semantic dependencies can be determined, for instance, by the features of verbs: transitive verbs have optionally two or three overt arguments (monotransitive or ditransitive). The morphological features of overt nouns and pronouns allow us to determine the syntactic dependencies of a) which nominals belong to a verb phrase and b) which syntactic role an NP

[26] Constraint Grammar is not the only available formalism suitable for parsing languages such as Cree. For instance, Universal Dependencies (http://universaldependencies.org/) have been used for manual annotation of texts in Arapaho, an Algonquian language spoken in Wyoming (e.g. Wagner et al., 2016). Like Constraint Grammar, Universal Dependencies do not require linear order to determine relationships between words.

takes (e.g., for a direct VTA, an obviative noun cannot be subject). Syntactic dependencies will also occur between nouns and determiners; the relative positions of the two (DEM N or N DEM) have been described as an NP and a predicating phrase respectively, though both are considered NPs in the CG parser.

The CG formalism for this parser uses the VISLCG-3 compiler (e.g. Bick & Didrikson, 2015; documentation can be found at http://visl.sdu.dk/cg3.html). This newer compiler includes various capabilities not available in earlier versions, such as the use of regular expressions in constraints, relationships between constraints, easier control over the scope of parameters, chunking heads and modifiers, and dependencies between objects and their complements or between anaphoric or discourse relationships. Of these, only regular expressions are currently used in the Plains Cree CG parser.

4.1. Disambiguation

Very little disambiguation is required for Plains Cree compared to languages such as Swedish, as there is considerably less homophony (see details below). A considerable degree of disambiguation is required due to ambiguity introduced by the morphological analyser, not the language itself: one version of the morphological analyser ignores vowel length in order to recognise as much as possible, as many speakers do not consistently mark vowel length and some printed works follow conventions different from those underlying the digitized corpus. When this version of the analyser is used on standard texts, simple constraints choose the analysis that matches the spelling. However, because the current input is the Gold Standard corpus, only correct analyses are included and so such constraints are redundant. When non-standard spelling is used, future constraints will be developed to begin to deal with these. Other common ambiguities include word forms that can be both particles and another word class, typically demonstratives or pronouns. In cases like these, weighted constraints would be the best approach, as these words are likely to be particles. However, even with translations available, researchers familiar with Cree cannot satisfactorily choose an option for disambiguation, and as such constraints cannot be written for all cases.

In the present iteration of the CG parser for Plains Cree, disambiguation between demonstratives and homophonous particles is achieved by their proximity to nouns or verbs with matching animacy and number features; this is a temporary solution as further research into these particles, especially *ôma*, is required. Ahenakew (1987) discusses and exemplifies several functions of *ôma*, and the examples encountered during manual disambiguation suggest that in conjunction with nouns, the demonstrative reading is very likely. The demonstrative *ôma*, among other pronouns, can also be used as the pronominal arguments of verbs, again when the number and animacy features correspond. In other cases, *ôma* is a frequent focus particle as well; in the current CG, the focus particle reading will be chosen when there is no verb or noun that *ôma* could otherwise be dependent upon, though this is neither the most accurate nor the most elegant solution.

Various other ambiguities occur that can be resolved by researchers on the basis of translation by fluent speakers, but these ambiguities do not affect the assignment of dependencies. In these cases, the analyses can either be collapsed to one analysis with a tag that represents the ambiguity (e.g. an obviative word can be considered both singular and plural), or a constraint can select the morphologically least complex form. The latter solution is employed for diminutive nouns and lexicalised verbal compounds. In the CG parser, diminutives and lexicalised verbal compounds are represented like the following:

(20) Diminutive
"<nêhiyâsisak>"
; "nêhiyaw" N AN Der/Dim N AN Pl[27]
"nêhiyâsis" N AN Pl

(21) Lexicalised verbal compounds
"<ê-kî-miyo-wîcêwak>"
; PV/e PV/miyo "wîcêwêw" V TA Cnj Prt 1Sg 3SgO
PV/e "miyo-wîcêwêw" V TA Cnj Prt 1Sg 3SgO

With the use of a translation, it is easy to select the preferred analysis. In these cases, both would be the second option: in (20), the translation is ‘young Crees’, which is a lexical entry in the dictionary (not ‘Crees of diminutive size’), and in (21), the translation hints that this is the second option, though both analyses can be translated as ‘I got along well with him’. The compiler gives the option displaying the dispreferred analyses for testing purposes, but removes them as possible analyses by means of the semicolon <;> at the beginning of the line. Similarly, possessed obviatives can be either singular or plural, as in (22).

(22) Obviative possessum
"<otawâsimisiwâwa>"
"awâsis" N AN Der/Dim N AN Sg Px3Pl Obv
"awâsis" N AN Der/Dim N AN Pl Px3Pl Obv

Again, it is easy for a researcher to select the second option on the basis of the translation: ‘their children’ or ‘their child’. For the diminutives and verbal compounds, the constraints select the morphologically least complex form that occurs in the lexicon. For obviative possessed nouns, disambiguation makes no difference to the assignment of dependencies, and so they can be simplified to one combined analysis in the preprocessing phase, giving a cohort as in (23).

(23) Obviative possessum, revised
"<otawâsimisiwâwa>"
"awâsis" N AN Der/Dim N AN Sg Pl Px3Pl Obv

These constraints make the assignment of dependencies more straightforward, but the morphosyntactic tags that dependency constraints need to refer to do not differ between the analyses, the disambiguation of such forms has no meaningful consequence in the parser. A further frequent ambiguity occurs when the preverbs *ka-* and *kâ-* are reduced to *k-* before vowel-

---

[27] Abbreviations for the morphological analyser and CG parser: N=noun, Der/Dim=derived diminutive, PV/=preverb, Prs=present, Prt=past, O=Object, Px=possessed by, Err/Orth=orthographic difference from analyser, Med=medial; @N=associated with a noun, @ACTOR=actor associated with a verb, @GOAL=goal associated with a verb, @PRED=verb, @OBL=oblique associated with a verb, @P=associated with a particle. Other abbreviations are familiar from Cree glossing.

initial morphemes. Both of these are preverbs on conjunct verbs, and so the features of word forms are otherwise identical, as in (24):

(24) "<k-êtohtêcik>"
PV/ka "itohtêw" V AI Cnj Prs 3Pl Err/Orth
PV/kaa "itohtêw" V AI Cnj Prs 3Pl Err/Orth

In some cases, the translation is sufficient for disambiguation by a researcher (*ka-* in cases of infinitive translations with 'to', or modality as in 'should, would'; *kâ-* with translations of 'when, where, who'). However, as this cannot be done with context-based constraints, and choosing one over the other does not affect the assignment of actors and goals, *kâ-* has been chosen as the preferred form on the basis of frequency.[28] A total of 67 disambiguation constraints are used in the present iteration of the parser; many constraints occur in pairs, contrasting values such as animate/inanimate, singular/plural, and to the left/right of the target.

Though not the focus of this paper, the effectiveness of the disambiguation constraints can be examined. The HP texts that have been manually coded for disambiguation and function assignment (discussed in more detail below), 544 analyses are manually removed, while the disambiguation constraints remove a total of 374 analyses in this same portion of the text. Of these, 335 are removed in both the manually-coded text and by the disambiguation constraints. Therefore, the recall rate for removal of analyses is 62% and the precision rate is 90%. Again, many of the problematic cases are those which cannot be determined by sentential context alone, and so lower rates are to be expected. For analyses that are not removed (i.e., treated as correct or preferred), 3241 are marked as such in the manually-coded text and 3202 remain after the disambiguation constraints have been applied; therefore, the recall rate for correctly preferred analyses is 99%. A closer look within the cohorts found in this portion of the texts is also possible. In Table 1 below, the number of Plains Cree word forms (i.e. excluding punctuation tokens) with certain numbers of analyses before and after disambiguation are given. When a correct analysis remains, this is indicated with a plus sign <+>; when a correct analysis is removed, this is indicated with a minus <->.

Table 1 Disambiguation results

Number of word forms	Analyses before disambiguation	Analyses after disambiguation	Accuracy	% of word forms
2,704	1	1	+	83.8
276	2	1	+	8.6
199	2	2	+	6.2
18	4	1	+	0.6
13	2	1	-	0.4
6	3	2	+	0.2
5	4	1	-	0.2
3	3	1	+	0.09
2	4	2	+	0.06
2	3	3	+	0.06

Of these word forms, those that have only one analysis both before and after disambiguation are assumed correct as they are drawn from the manually-verified morphological

[28] In the HP texts, the preverb *kâ-* occurs unambiguously in 91 tokens and *ka-* in 35 tokens; the ambiguous reduced *k-* occurs in 37 tokens, but *ka-* can be confidently selected as the preferred form for only one token.

Gold Standard corpus.[29] There are 529 word forms with 2 or more analyses before disambiguation; 297 of these have one correct analysis after disambiguation. This gives a recall rate of 56% for the selection of the one correct analysis. After disambiguation, 315 forms are reduced to one analysis; the 297 with the correct analysis remaining give a precision rate of 94%.[30] There are 209 word forms with more than one analysis remaining after disambiguation, 6% of the 3,226 total word forms presented here. However, in all cases where more than one analysis remains, the correct analysis has not been removed.

Applied to the whole corpus, the disambiguation constraints appear to reduce ambiguity to a comparable degree. The counts for both unique types and individual tokens for different numbers of possible analyses are compared before and after disambiguation in Table 2 and Table 3 respectively. First, these tables demonstrate how little ambiguity is found in Plains Cree; the majority of forms receive just one analysis and forms with more than two analyses are rare. For those that receive multiple analyses (before disambiguation), the most common are unsurprisingly those that can be both pronouns and particles or both animate and inanimate, which often cannot be disambiguated by sentential context: *ôma* (n = 1,254), *anima* (n = 1,022), *êkoni* (n = 370), *anihi* (n = 267), *ôhi* (n = 210). Of these, *ôma* accounts for 21% of the total ambiguous forms in the corpus; these five most frequent forms account for 50% of the total ambiguity.

Table 2 Type Ambiguity

	Before disambiguation		After disambiguation	
# of analyses	Number of types	% of types	Number of types	% of types
0[31]	2,255	11.7	2,255	11.7
1	14,563	75.8	16,183	84.0
2	2,205	11.5	794	4.1
3	109	0.6	31	0.2
4	88	0.46	7	0.04
5	3	0.02	0	0.00
6	1	0.005	0	0.00

Table 3 Token Ambiguity

	Before disambiguation		After disambiguation	
# of analyses	Number of tokens	Proportion	Number of tokens	Proportion
0	3,293	3.9	3,293	3.9
1	65,046	77.2	75,134	89.2
2	12,933	15.3	5,739	6.8
3	1,014	1.2	103	0.1
4	1,980	2.3	7	0.008
5	9	0.01	0	0
6	1	0.001	0	0

The disambiguation constraints, though not complete, do reduce the ambiguity found in the corpus considerably. For example, before disambiguation there are 2,205 types with 2

[29] Though single analyses here are understood as correct, a single analysis will never be removed by a CG parser. In texts where analyses have not been verified, a word form with only one analysis cannot be guaranteed correct.
[30] While lower than the more general rates given above, these still suggest that the disambiguation constraints benefit the analysis overall.
[31] Zero represents forms with no analysis, which include unrecognized Cree forms and other forms that are not yet explicitly tagged appropriately non-Cree, such as English words or fragments.

analyses and after there are 794 types, and before disambiguation there are 1,014 tokens with 3 analyses and after there are 103 tokens. Once ambiguous forms have been disambiguated to the extent possible, constraints for identifying dependencies are then applied.

4.2. Dependencies

Several main dependencies are implemented in the Plains Cree CG parser at this time. However, the key relationships that are of interest in the present work are the relationships between verbs and their lexicalised arguments and the internal structure of noun phrases: the occurrence of pronouns, demonstratives, and nouns. Demonstratives such as *ôma* are disambiguated by their proximity to nouns, and these same relationships are used to identify the dependency relations as well. The constraints shown in (25) mark the dependency @N ("dependent on a noun") on an animate singular or plural demonstrative (the TARGET) when the word form in question has a noun with matching features (animate plus singular or plural) immediately to the left or right. The angle brackets > and < indicate the direction of the dependency, much like the arrows of dependency syntax. Barriers determine the range within which the constraint will apply; CLB represents a predetermined list of clause boundaries, introduced at the beginning of §4. An example from the corpus is given in (26).

(25) Demonstrative and noun dependencies

a. MAP:DemNANSgR @<N TARGET Dem + AN + Sg IF (NOT -1 Obv)(-1 N + AN + Sg BARRIER CLB) ;
b. MAP:DemNANSgL @N> TARGET Dem + AN + Sg IF (NOT 1 Obv)(1 N + AN + Sg BARRIER CLB) ;
c. MAP:DemNANPlR @<N TARGET Dem + AN + Pl IF (NOT -1 Obv)(-1 N + AN + Pl BARRIER CLB) ;
d. MAP:DemNANPlL @N> TARGET Dem + AN + Pl IF (NOT 1 Obv)(1 N + AN + Pl BARRIER CLB) ;

(26) "<nisis>"
"-sis" N AN Sg Px1Sg
"<ana>"
"awa" Pron Dem Med AN Sg @<N

The constraints that map dependencies between arguments and their governing verbs are considerably more complex than those for demonstratives and nouns, though the complex morphology and resulting morphosyntactic tags make most relationships plain. The following set of constraints identifies goals of VTAs, both obviative and proximate. In (27)a., a singular animate noun is marked as a goal if anywhere to the left in the clause there is a transitive animate verb with a tag for a singular obviative goal and there are no intervening verbs.[32] In (27)b., the constraint requires the same context but to the right of the noun. The constraints in (27)c. and (27)d. perform the same identification of goals, but this time with plural animate nouns. An animate noun is the goal if it is not obviative (NOT 0 Obv) and the verb's tag string includes the appropriate third person object tag, either plural or singular. Then, the constraints in (27)e. and

[32] Introducing a noun as a barrier may solve some problems, but introduces others. Further refinement of these rules is required.

(27)f. do the same for obviative nouns, as only animate nouns can be marked with the tag Obv, these constraints can be simpler.

(27) Goals of TA verbs
   a. MAP:TAGOAL3R @<GOAL TARGET N + AN + Sg IF (NOT 0 Loc)(NOT 0 Obv)(*-1 TA + 3SgO BARRIER V OR CLB) ;
   b. MAP:TAGOAL3L @GOAL> TARGET N + AN + Sg IF (NOT 0 Loc)(NOT 0 Obv)(*1 TA + 3SgO BARRIER V OR CLB) ;

   c. MAP:TAGOAL3R @<GOAL TARGET N + AN + Pl IF (NOT 0 Loc)(NOT 0 Obv)(*-1 TA + 3PlO BARRIER V OR CLB) ;
   d. MAP:TAGOAL3L @GOAL> TARGET N + AN + Pl IF (NOT 0 Loc)(NOT 0 Obv)(*1 TA + 3PlO BARRIER V OR CLB) ;

   e. MAP:TAGOALObvR @<GOAL TARGET N + Obv IF (NOT 0 Loc)(*-1 TA + 4Sg/PlO BARRIER V OR CLB) ;
   f. MAP:TAGOALObvL @GOAL> TARGET N + Obv IF (NOT 0 Loc)(*1 TA + 4Sg/PlO BARRIER V OR CLB) ;

Lexicalised actors are similarly assigned @{<}ACTOR{>} functions; when the verbs are marked for third person actors, nouns matching in number and animacy are assigned as such. Beyond nouns, pronouns can also function as actors in Plains Cree sentences. These can be both personal pronouns (*niya* '1SG', *kiya* '2SG', etc.) and the demonstrative pronouns (*ôma, awa,* and their paradigms including inflection for number, proximity, and obviation). Like nouns, the relationships between pronouns and verbs can be identified with these features: first person pronouns as arguments of verbs with first person features, animate pronouns as arguments of verbs with third person features, etc. Pronouns are not arguments in the cases where they are dependent on nouns, and this must be reflected in the constraints to avoid overapplication of the argument functions.

Once arguments are identified by means of the class features of verbs, there are still other nouns and pronouns that may need to be accounted for. These nouns occur often for VAI and VTA clauses, where additional nominal elements (usually inanimate, but not always), are associated with the verb but cannot be an "argument" in terms of the argument structure described for the verb class. These may function as objects of VAIs (e.g., *minihkwêw* 's/he drinks' is structurally a VAI, but in most cases requires an object to be drunk) or the indirect objects of ditransitive VTAs. In the current version of the CG parser, any "leftover" nouns within a clause (as delineated clause boundaries, including commas) is identified as an oblique (@OBL) associated with the nearest verb. Oblique constraints are implemented in the current CG, but are not referenced in the analysis and discussion below.

The current CG parser includes 105 dependency mapping constraints. As for disambiguation constraints, these frequently occur in groups of feature pairs, with otherwise identical constraints mapping dependencies between nouns and demonstratives and between nominal arguments and verbs, with all permutations of animacy and number, as well as obviation, occurring for each constraint type to both the left and the right of the target noun or demonstrative. Thus, 10 constraints are used for relationships between nouns and demonstratives, and 72 constraints are devoted entirely to the assignment of actor and goal functions. A further 10 constraints are used to identify the pronoun *êkoni* as dependent on a noun, as it is not tagged like demonstrative pronouns and does not inflect with the same features. Four constraints mark verbs as predicates. The remaining constraints, though not relevant to the results presented in this paper, briefly assign oblique functions and recognise locatives as dependent on particles.[33]

5. Results and discussion

5.1. Model coverage

To evaluate the coverage of the CG parser, a subset of the Plains Cree morphological Gold Standard corpus, the HP texts (Vandall & Douquette, 1987), was manually coded for the assignment of actors, goals, and relationships between nouns and demonstrative pronouns.[34] This text collection includes approximately 3,200 tokens, or roughly 4.5% of the Plains Cree tokens in the corpus (excluding punctuation). Of primary interest in this paper is the coverage of the model with respect to the assignment of actors and goals; the current model has both a recall and precision rate of 92%. Where mismatches of actor and goal assignment occur for individual nouns (n = 15), these are nouns that have the appropriate features to be the arguments of different verbs on either side of them, though the parser has selected an option different from my manual coding. In situations where I have identified an actor or a goal and the CG has not (n = 9), further refinement of constraints, particularly where pronouns and obviative nouns are concerned, is required. Where the CG has assigned an actor or goal reading and I have not (n = 15), the incorrectly assigned actors (n = 5) occur because the CG has assigned an inanimate actor, e.g. *ôma*, to an impersonal VII such as *kîsikâw* 'it is day'. The incorrectly assigned goals are all cases of *ôma* being misidentified as a pronominal VTI goal rather than a particle.[35] As *ôma* is generally a problematic case for correct disambiguation, these situations cannot be solved only with context and morphosyntactic features and instead require the addition of semantic information.

This version of the CG was then applied to the full Plains Cree corpus. The corpus is comprised of approximately 91,000 tokens (both Cree and other languages, such as English and French, as well as punctuation) and includes texts that can be found in several published volumes: Ahenakew (2000), Bear et al. (1992), Kâ-Nîpitêhtêw (1998), Masuskapoe (2010), Minde (1997), Vandall and Douquette (1987), Whitecalf (1993). Over 72,000 of these words are Plains Cree that receive an analysis in the Gold Standard. These texts are transcribed from narratives and conversations recorded in the 1980s and 1990s, spoken by elderly fluent speakers of Plains Cree from Alberta and Saskatchewan. Of these texts, one (Ahenakew, 2000) was not included in the Gold Standard, instead set aside for further testing of the morphological model. However, 95% of the word forms in this text are attested elsewhere in the corpus and were therefore included in the Gold Standard analyses. The full corpus has been analysed using the above-described Constraint Grammar parser and results are reported in the following subsections.

---

[33] See Appendix A for more details.

[34] This collection of texts was also used by Wolvengrey (2011) to demonstrate the various iterations of transitive verbs and their arguments that occur.

[35] Lexicosemantic coding will be a useful tool for refining this in future. See Appendix A for more details. Furthermore, preprocessing for particle phrases will also be useful for the disambiguation of *ôma* as a particle.

5.2. Phrase orders

5.2.1. Clause patterns in the corpus

When applied to the corpus, the CG can be used to determine the overall number of clauses (as delineated by clause-boundary punctuation, e.g. commas and periods) in the corpus, and the patterns which these clauses exhibit. Though not the focus of this paper, these clauses demonstrate the types that are seen in Plains Cree. There are 5,238 unique clause patterns found in 31,433 total clause tokens. The most common (occurring in one percent or more of clause tokens) are given in Table 4. For each pattern, the percent of the total 31,433 clauses is included. These clauses include word forms to which functions have been assigned (e.g., PRED-TA is a verb) and those that are not yet included in the constraints, represented only by their part of speech tag (e.g. IPC is a particle). The most frequent clauses are those that contain no Cree material (0); these include metadata such as speaker identification and comments on the text, as well as punctuation that was separated by preprocessing, English and French words, and numerals. Note that overt arguments are not represented these most frequent patterns, though arguments are the focus of the following subsections.

Table 4 Overall clause patterns

Frequency	Clause patterns			% of 31,433
3423	0			10.9
3225	Ipc			10.3
1723	@PRED-AI			5.5
1003	Ipc	Ipc		3.2
906	N			2.9
755	@PRED-TA			2.4
686	Ipc	@PRED-AI		2.2
540	Ipc	Ipc	@PRED-AI	1.7
534	Ipc	Ipc	Ipc[36]	1.7
364	@PRED-TI			1.2
362	@PRED-AI	Ipc		1.2
332	Pron			1.1

5.2.2. Overall verb and argument orders

These phrases can also be simplified to examine only the counts for various configurations of verbs and noun phrases within clause boundaries (as marked by commas, periods, and similar punctuation). The verb is the centre of the clause around which word orders can be determined. There are 1,851 clauses containing VIIs, 8,604 containing VAIs, 3,870 containing VTIs, and 5,409 containing VTAs, for a total of 19,734 clauses containing verbs (63% of the 31,433 in the corpus).[37] In Table 5 below, the twenty-five most frequent

---

[36] Though the possibility of three or four particles constituting a single clause (that is marked out by punctuation) may suggest an error, such clauses do indeed occur, as in (i).

(i) *êkw ânim êkota êkwa*

êkwa	anima	êkota	êkwa
and/then	FOC	there	and/then

'and then there' (Ahenakew, 2000, pp. 32-3)

The implementation of subclasses of particles and the inclusion of particle phrases in the future will be useful for determining the internal patterns of phrases consisting of multiple particles.

[37] These counts are considerably larger than those for verb class tokens presented in Schmirler & Harrigan (2016). There are two main reasons for these discrepancies: one, the Gold Standard has improved considerably since that

configurations of tagged verbs and arguments are presented.[38] The percentages of the 19,734 total verbal clauses are also included.

Table 5 Phrase order frequencies

Frequency	Verb class and arguments			% of 19,734
4865	@PRED-AI			24.7
2128	@PRED-TA			10.8
1432	@PRED-TI			7.3
801	@PRED-II			4.1
539	@PRED-TA	@<GOAL		2.7
521	@ACTOR>	@PRED-AI		2.6
503	@PRED-AI	@<ACTOR		2.5
450	@GOAL>	@PRED-TI		2.3
386	@GOAL>	@PRED-TA		2.0
324	@PRED-TI	@<GOAL		1.6
242	@PRED-AI	@PRED-AI		1.2
211	@ACTOR>	@PRED-II		1.1
173	@ACTOR>	@PRED-TA		0.9
162	@PRED-AI	@N>	@<ACTOR	0.8
159	@PRED-II	@<ACTOR		0.8
148	@PRED-TA	@<ACTOR		0.7
132	@PRED-TA	@N>	@<GOAL	0.7
116	@PRED-TA	@PRED-AI		0.6
114	@PRED-TI	@PRED-AI		0.6
111	@ACTOR>	@PRED-TI		0.6
103	@PRED-TI	@<ACTOR		0.5
76	@PRED-TI	@N>	@<GOAL	0.4
72	@PRED-II	@PRED-AI		0.4
71	@PRED-TA	@PRED-TA		0.4
64	@PRED-AI	@PRED-TA		0.3

This table includes single verbs without overt arguments; for all verb classes, these are by far the most frequent occurrences. A single VAI with no arguments, for example, occurs in almost ¼ of the total clauses containing verbs in the corpus. Similarly, where there are two verbs in a single clause, these occur most frequently without overt arguments. These patterns are unsurprising given not only the extensive person morphology on Cree verbs, but also the narrative and dialogue nature of the texts: speakers are often relating stories about themselves or speaking to each other, making first and second person forms quite common. Similarly, as expected in any discourse, once a referent has been introduced it is unlikely they will be referred to explicitly in every following clause that refers to them, as discussed above (e.g. Wolvengrey, 2011). Where arguments do occur, we begin to see patterns discussed in smaller, qualitative investigations confirmed on a larger scale: the most frequent occurrence is that of a VTA followed by an overt goal. This order might be expected of situations where a new referent, not a topic or in focus, is introduced to the discourse. As this referent is not pragmatically marked, it does not occur before the verb (cf. Mühlbauer's basic VSO order). Next, with very similar frequency counts, are VAIs with an actor preceding (n = 521) and following (n = 503). VAIs are

---

analysis and two, the CG parser has introduced some discrepancies with multiple verbs that could not be contextually disambiguated. Manual scrutiny and further development and testing of the CG are required. To minimise the effects of such idiosyncrasies, only more frequent orders are presented here.

[38] More comprehensive tables are included in Appendix B.

overall the most common verb type, and as they cannot take goals, these counts are in line with expectations.

Following that, we begin to see VTIs and VTAs with goals preceding them. Following Wolvengrey (2011) and others, these would be placed in a topic or focus position preceding the verb. While goals follow VTAs more often than they precede them, this is not the case for VTIs. Topic and focus in the discourse may also account for these differences. Similarly, actors preceding VTAs (n = 173) are next most frequent.[39] Actors following VTAs are less common still; this may be due to the unlikelihood of a goal being specified in the discourse but the actor less topical or unspecified. Investigations of topic and focus will require, minimally, consideration of verbal marking, particularly inverse marking on VTAs, and the internal structure of overt NP arguments: whether they contain only a noun, only a pronoun, or a demonstrative and a noun, and whether they are proximate or obviative, as well as the occurrence of non-lexicalised arguments. Pronouns and proximate NPs, as inherently more topical, may be more likely to precede verbs than complex or obviative NPs.

Further investigations into topic and focus will require several steps. First, the above-mentioned features such as the complexity of the NP and obviation need to be considered. One would also note when an argument occurs overtly in a previous clause, and how many referents, if any, intervene. Semantic coding for certain resumptive pronoun constructions would also allow for understanding how referents are coded. The first of these is a particle phrase: *êwako* occurs before demonstratives such as *ôma*, creating a phrase that can be translated as ‘this (previously mentioned) one’. Similarly, *êkoni* (inanimate) and *êkonik* (animate) are single pronouns that can be similarly translated, but they themselves inflect more like pronouns with at least some possible combinations of animacy, number, and obviation; they can also be followed by pronouns, nouns, or demonstrative-noun phrases of the same features. These resumptive pronominal elements are said to link to the topic of preceding discourse; they are more common in clause-initial positions, before the verb, but may also occur post-verbally, perhaps in a final position (Wolvengrey, 2011, pp. 207, 227-40). The capabilities of VISLCG-3 in terms of anaphora and discourse relationships may be valuable tools in quantitative investigations of topic and focus.

5.2.3. Transitive animate phrase orders

As VTAs can take up to two morphologically marked lexicalised arguments, they provide a more interesting array of possible phrase orders to examine. Additionally, Wolvengrey’s (2011) analysis of the VTA argument orders in Vandall & Douquette (1987) offers a starting point for comparison between a smaller, qualitative study and a quantitative investigation of word order in a larger, though still relatively small, corpus of Plains Cree. In Table 6 below are the counts for configurations of arguments and transitive animate verbs in the corpus; here, only clauses containing one verb, and only those orders that occur at least five times, are included. There is a total of 4,551 clauses containing one VTA in the corpus, or 23% of all clauses containing verbs.

[39] At this point, I do not comment on VII actors due to the frequent misidentification of the particle *ôma* as the actor of an impersonal VII, as this issue will have likely led to quite inaccurate counts. However, the counts are still included in the above table as a representation of how the CG identifies VII actors using only person, number, and animacy information.

Table 6 VTA phrase order frequencies

Frequency	VTA phrase orders				% of 4,551
2128	@PRED-TA				46.8
539	@PRED-TA	@<GOAL			11.8
386	@GOAL>	@PRED-TA			8.5
173	@ACTOR>	@PRED-TA			3.8
148	@PRED-TA	@<ACTOR			3.3
132	@PRED-TA	@N>	@<GOAL		2.9
54	@PRED-TA	@N>	@<ACTOR		1.2
35	@ACTOR>	@PRED-TA	@<GOAL		0.8
28	@GOAL>	@PRED-TA	@<GOAL		0.6
23	@N>	@GOAL>	@PRED-TA		0.5
22	@PRED-TA	@<GOAL	@<GOAL		0.5
21	@ACTOR>	@PRED-TA	@<ACTOR		0.5
20	@GOAL>	@GOAL>	@PRED-TA		0.4
17	@ACTOR>	@GOAL>	@PRED-TA		0.4
17	@N>	@ACTOR>	@PRED-TA		0.4
13	@GOAL>	@<N	@PRED-TA		0.3
12	@PRED-TA	@<ACTOR	@<GOAL		0.3
12	@GOAL>	@PRED-TA	@<ACTOR		0.3
11	@GOAL>	@PRED-TA	@N>	@<GOAL	0.2
11	@GOAL>	@ACTOR>	@PRED-TA		0.2
9	@ACTOR>	@ACTOR>	@PRED-TA		0.2
9	@PRED-TA	@<ACTOR	@<ACTOR		0.2
8	@ACTOR>	@<N	@PRED-TA		0.2
6	@ACTOR>	@PRED-TA	@N>	@<GOAL	0.1
5	@PRED-TA	@<GOAL	@<N		0.1
5	@GOAL>	@GOAL>	@PRED-TA	@<GOAL	0.1
5	@PRED-TA	@<GOAL	@<ACTOR		0.1

Instances where a verb has its actor or goal labelled on both sides can be the result of several phenomena. One common pattern is sentences such as the following, where two demonstrative pronouns with the appropriate features are found on either side of the verb, but one of them is adjacent to the noun; similarly, there may simply be two pronouns on either side of the noun. Such patterns are exemplified in (28) and (29). Where two actors or goals precede or follow an argument, these are usually the result of two coordinated NPs, or a noun plus a personal pronoun, which are not yet linked in the CG. Others are errors introduced by the formatting of the text; there are situations where both a demonstrative and a noun are identified as the same argument, but cannot be linked because of intervening punctuation that does not mark a clause boundary; square brackets used to add additional characters and clarify the text are seen frequently. Finally, some of these instances are introduced by the CG itself, which cannot yet distinguish, for example, between direct and indirect objects.

(28) Multiple goals (Vandall & Douquette, 1987, pp. 72-3)

iyisâc	awa	ni-	atamiskaw-	-â	-w	awa	kisêyiniw
IPC	DEM.AN.SG	1	VTA	DIR	3SG	DEM.AN.SG	NA
reluctantly	this	I shook hands with him				this	old man
	@GOAL>	@PRED-TA		@N>		@<GOAL	

‘I reluctantly shook hands with this old man’

(29) Multiple actors (Bear et al., 1992, pp. 130-1)

ahpô êtikwê	awa	ê-	wî-	kiskinohtah-	-iko	-yâhk	awa
IPC	DEM.AN.SG	PV	PV	VTA	INV	1PL	DEM.AN.SG
maybe	this	he will show us the way					this
	@ACTOR>	@PRED-TA					@<ACTOR

‘maybe it [the owl] is going to show us the way’

The first several counts in Table 6 are found in Table 5 above and discussed there. As patterns become less frequent, there is a general trend away from the proposed basic VSO order: arguments before the verb, goals before actors. Similarly, more complex noun phrases begin to occur; again, topic and focus may be at play, but closer scrutiny is required. While more in-depth investigation into the internal structures of the nouns and verbs in the above orders would shed light on aspects of topic and focus, the above counts can also be simplified to the order of the verb, actor, and goal, as in Table 7. Instances where the same argument occurs on each side of the verb are excluded; cases where the same argument occurs twice on one side of the verb are treated as though they were a single argument. In the fourth column, the percentages found for monotransitive VTA phrases in HP texts by Wolvengrey (2011, p. 202) are also included. The HP texts represent approximately 3,200 (4.5%) of the over 72,000 words in the Plains Cree corpus. These 286 VTA clauses represent 6% of the 4,551 VTA clauses in the corpus.

Table 7 VTA phrase order comparison

	Plains Cree corpus		HP texts (Wolvengrey, 2011)	
Phrase order	Frequency	Percent	Frequency	Percent
V	2128	56.2	88	30.8
V GOAL	698	18.4	85	29.7
GOAL V	442	11.7	54	18.9
V ACTOR	211	5.6	15	5.2
ACTOR V	207	5.5	11	3.8
ACTOR V GOAL	41	1.1	8	2.8
ACTOR GOAL V	17	0.5	6	2.1
V ACTOR GOAL	12	0.3	3	1.0
GOAL V ACTOR	12	0.3	11	3.8
GOAL ACTOR V	11	0.3	3	1.0
V GOAL ACTOR	5	0.1	2	0.7
Total	3784	100.0	286	100.0

Between just Vandall and Douquette (1987) and the entire corpus, the percentages of phrase orders demonstrate similar patterns: a single verb is most common and V GOAL ACTOR is the least frequent. In between, the phrase order frequencies descend in nearly identical orders. The discrepancies may be due to inherent differences in the styles of texts in the collection; Vandall and Douquette (1987) is a series of shorter narratives, while other texts such as Bear et al. (1992), Kâ-Nîpitêhtêw (1998), and Masuskapoe (2010) include longer narratives and dialogues, where there may be more opportunity for non-lexicalised arguments as more verbs refer to the same topical participant. Additionally, Vandall and Douquette (1987) speak mostly about history and relate others’ stories, so there are many third person verbs. On the other hand, texts like Bear et al. (1992) involve more stories about the speaker’s own lives, and dialogues where speakers directly address each other, so there are likely to be a greater proportion of first and second person verbs. As for beginning investigations into topic and focus, the characteristics of VTAs in individual texts may also be an area for understanding phrase order in Plains Cree.

Nearly half of the attested VTA phrase orders each occur in less than 1% of phrases, including the purported basic VSO (V ACTOR GOAL) at 0.3%. However, when all verb-initial phrases are combined, this percentage increases considerably: these comprise 81% of the above phrases. While such a frequency may motivate VSO as a proposed basic word order, a claim of word order based on frequency including phrases where no lexicalised arguments occur would be a questionable choice. However, basic word order and most frequent word order need not correlate (cf. psycholinguistic evidence for Mayan, Koizumi et al., 2010). Further investigation is required to determine what, if any, order is basic in Plains Cree.

To minimise the effects of narrative type, e.g. history versus personal narrative, one step is to only examine VTAs with non-local arguments (third person proximate and obviative)—that is, verbs that can take two nouns as their arguments, not just pronouns. The total number of both non-local direct and inverse phrases in the corpus is 1202, 6% of the total clauses containing verbs and 26% of clauses containing VTAs; direct VTAs represent 82% of the non-local VTA phrases and the inverse phrases, 18%. The phrase orders of direct VTAs—where proximate (3) persons act on obviative (3′) persons—are presented in Table 8; only phrase orders containing one VTA that occur five or more times are included. Percentages of the total 985 non-local direct VTAs are given. For these clauses containing direct verbs, patterns differ little from those found for VTAs overall in the corpus. Lexicalised goals are more common than actors, arguments tend to occur after the verb more often than before it, and complex NP arguments are less frequent than those that contain only a noun or a pronoun. Again, a better understanding will be gleaned once part of speech and inflectional information are considered in the future. However, looking at direct and inverse verbs separately may begin to demonstrate how topicality affects lexicalisation and word order. Here, the obviative goals are the most common lexicalised arguments; when both actors and goals are lexicalised, it is more common for the more topical actors to occur before the obviative goals.

Table 8 VTA phrase order 3 > 3′

Frequency	Phrase order			% of 985
303	@PRED-TA			30.8
189	@PRED-TA	@<GOAL		19.2
133	@GOAL>	@PRED-TA		13.5
47	@ACTOR>	@PRED-TA		4.8
46	@PRED-TA	@N>	@<GOAL	4.7
37	@PRED-TA	@<ACTOR		3.8
29	@ACTOR>	@PRED-TA	@<GOAL	2.9
15	@PRED-TA	@N>	@<ACTOR	1.5
10	@ACTOR>	@GOAL>	@PRED-TA	1.0
8	@GOAL>	@PRED-TA	@<GOAL	0.8
8	@GOAL>	@PRED-TA	@<ACTOR	0.8
8	@GOAL>	@ACTOR>	@PRED-TA	0.8
8	@ACTOR>	@PRED-TA	@<ACTOR	0.8
7	@PRED-TA	@<ACTOR	@<GOAL	0.7
6	@PRED-TA	@<GOAL	@<GOAL	0.6
5	@ACTOR>	@ACTOR>	@PRED-TA	0.5

While non-local direct VTAs differ little from the overall patterns found for VTAs across the corpus, inverse VTAs demonstrate different, but expected, patterns. In Table 9, the phrase orders for non-local inverse VTAs that occur two or more times in the corpus are presented. Percentages of the total 217 non-local inverse VTAs are also included. First, the number of phrases is considerably different; where direct VTAs with no lexicalised arguments occur 303 times, inverse VTAs with no arguments occur only 87 times. Next, actors occur before goals in the same position: V ACTOR before V GOAL and ACTOR V before GOAL V, unlike both direct non-local VTAs and VTAs in general. Finally, complex NPs are seen for actors, but not for goals. In contrast to the patterns seen for the direct verbs above, when both arguments are lexicalised for inverse clauses, it is the goals (now the proximate, more topical arguments) that occur first.

Table 9 VTA phrase order 3′ > 3

Frequency	Phrase order				% of 217
87	@PRED-TA				40.1
21	@PRED-TA	@<ACTOR			9.7
15	@PRED-TA	@<GOAL			6.9
15	@ACTOR>	@PRED-TA			6.9
7	@GOAL>	@PRED-TA			3.2
4	@PRED-TA	@N>	@<ACTOR		1.8
2	@GOAL>	@PRED-TA	@N>	@<ACTOR	0.9
2	@GOAL>	@PRED-TA	@<ACTOR		0.9
2	@ACTOR>	@PRED-TA	@<ACTOR		0.9

These patterns demonstrate the differences between proximate (more topical) and obviative (less topical) actors: proximate actors, as the more topical participants, are lexicalised less frequently than the less topical obviative actors. When both are lexicalised, the more topical tends to occur first, regardless of its semantic role. These counts may also demonstrate a link between complex NPs and topicality, such that less topical participants may occur with demonstratives more often. However, further investigation is required to determine how reliable this tendency may be for establishing topicality or otherwise. Furthermore, as the frequency counts here are quite small, this may not be an accurate picture of topicality and word order.

6. Conclusion

Using only the morphosyntactic features output by a morphological model for Plains Cree, a Constraint Grammar parser can successfully begin to demonstrate patterns in the order of verbs and arguments in a Plains Cree corpus. Without any semantic information, the parser can successfully identify 92% of actors and goals in a manually-coded portion of the texts; the majority of the errors involve issues that would be solved by the addition of semantic coding. Semantic classifications within verb classes and among particles are the next steps in improving these problematic cases.

The patterns seen in the frequency counts supplied by the CG analysis are unsurprising compared to patterns described in smaller qualitative investigations: all verb classes occur most often without any lexicalised arguments, goals tend to be more frequent than actors, but when both occur, actors tend to occur first. More complex noun phrases seem to correspond with (lack of) topicality, similarly occurring less frequently. Within VTAs, direct and inverse verbs display different patterns in accordance with the relative topicality of their participants: topicality appears to have more bearing on the lexicalisation and order of the arguments than semantic roles. Across all verbs, further investigations into the features of individual noun phrases and the verbs with which they occur are required to truly begin to understand the relationships between topic, focus, and word order.

This paper has described the features of Plains Cree morphosyntax and the development of a Constraint Grammar parser required to assign actors and goals in a Plains Cree corpus. These labels can be implemented in a web-based corpus interface (Schmirler et al., 2017), and the parser can form the basis of the first iteration of a Plains Cree grammar checker, an accompaniment to a spell checker based on the morphological model. Both tools can also be integrated into intelligent Computer-Assisted Language Learning (iCALL) programs for Plains Cree (e.g. Bontogon, 2016). As the first computational tool of its kind for Plains Cree, a CG parser opens avenues for continued large-scale quantitative investigations of the syntax of Plains Cree and related dialects and languages. Further development of the parser will increase its accuracy and its functions, allowing for full dependency analyses of Plains Cree sentences and a better quantitative understanding of word order and its interactions with topic and focus in Plains Cree.

7. References

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*. H.C. Wolfart (Ed.). Winnipeg: University of Manitoba Press.

Ahenakew, F. (1987). Cree language structure: A Cree approach. Winnipeg: Pemmican Publications.

Antonsen, L., Huhmarniemi, S., & Trosterud, T. (2009). Constraint grammar in dialogue systems. In *Proceedings of the 17th Nordic Conference of Computational Linguistics. NEALT Proceeding Series* (Vol. 8, pp. 13-21).

Arppe, A. (2000). Developing a grammar checker for Swedish. In *Proceedings of NODALIDA* (Vol. 99, pp. 13-27).

Bear, G., Fraser, M., Calliou, I., Wells, M., Lafond, A., & Longneck, R. (1992). *kôhkominawak otâcimowiniwâwa / Our Grandmothers' Lives as Told in Their Own Words*. F. Ahenakew & H.C. Wolfart (Eds.). Regina: Canadian Plains Research Center.

Bick, E., & Didriksen, T. (2015, May). CG-3—Beyond Classical Constraint Grammar. In *Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania* (No. 109, pp. 31-39). Linköping University Electronic Press.

Birn, J. (2000). Detecting Grammar Errors with Lingsoft's Swedish grammar checker. In *Proceedings of NODALIDA* (Vol. 99, pp. 28-40).

Blain, E. (1997). WH-constructions in Nêhiyawêwin. Ph.D. thesis, University of British Columbia.

Blain, E. M. (1999). Nêhiyawêwin nominal clauses. In *Papers of the 30th Algonquian Conference*, D.H. Pentland (Ed.)., 12-27. Winnipeg: University of Manitoba.

Bloomfield, L. (1934). *Plains Cree texts*. American Ethnological Society Publications 16. New York. Reprinted 1974, New York: AMS Press.

Bontogon, M. A. (2016). "Evaluating nêhiyawêtân: A computer assisted language learning (CALL) application for Plains Cree." (Honours thesis). Retrieved from University of Alberta Electronic and Research Archives. (doi:10.7939/R3VD6P81C).

Branigan, P., & MacKenzie, M. (1999). A double-object constraint in Innu-aimûn. In *Papers of the 30th Algonquian Conference*, D.H. Pentland (Ed.)., 12-27. Winnipeg: University of Manitoba.

Cook, C. (2014). *The Clause-typing system of Plains Cree*. Oxford: Oxford University Press.

Dahlstrom, A. (1991). *Plains Cree morphosyntax*. New York: Garland.

Dahlstrom, A. (1995). *Topic, focus and other word order problems in Algonquian*. Winnipeg: Voices of Rupert's Land.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language 67(3)*, 547-619.

Graffi, G. (2001). *200 Years of Syntax: A critical survey* (Vol. 98). John Benjamins Publishing.

Harrigan, A. G., Schmirler, K., Arppe, A., Antonsen, L., Trosterud, T., & Wolvengrey, A. (forthcoming). Learning from the Computational Modelling of Plains Cree Verbs: Analysis and Generation Using Finite State Transduction. *Morphology*.

Hengeveld, K., & Mackenzie, J. L. (2008). *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford University Press.

Hirose, T. (2003). *Origins of Predicates: Evidence from Plains Cree*. Routledge.

Junker, M. (2004). Focus, obviation, and word order in East Cree. *Lingua*, *114*(3):345-65. doi:10.1016/S0024-3841(03)00027-5.

Karlsson, F. (1995a). Designing a parser for unrestricted text. In F. Karlsson, A. Voutilainen, J. Heikkilae, & A. Anttila (Eds.), *Constraint Grammar: a language-independent system for parsing unrestricted text* (Vol. 4) (pp. 1-40). Walter de Gruyter.

Karlsson, F. (1995b). The formalism and environment of Constraint Grammar Parsing. In F. Karlsson, A. Voutilainen, J. Heikkilae, & A. Anttila (Eds.), *Constraint Grammar: a language-independent system for parsing unrestricted text* (Vol. 4) (pp. 41-88). Walter de Gruyter.

Kâ-Nîpitêhtêw, J. (1998). *ana kâ-pimwêwêhahk okakêskihkêmowina / The Counselling Speeches of Jim Kâ-Nîpitêhtêw*. F. Ahenakew & H.C. Wolfart (Eds.). Winnipeg: University of Manitoba Press.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2015). Plains Cree. In *Ethnologue: Languages of the World, 18th edition*. Dallas, Texas: SIL International. Online version. http://www.ethnologue.com. Accessed November 15, 2015.

Masuskapoe, C. (2010). *piko kîkway ê-nakacihtât: kêkêk otâcimowina ê-nêhiyawastêki*. H.C. Wolfart and F. Ahenakew (Eds.). Winnipeg: Algonquian and Iroquoian Linguistics.

Mel′čuk, I. A. (1988). *Dependency syntax: theory and practice*. State University of New York Press: Albany.

Minde, E. (1997). *kwayask ê-kî-pê-kiskinowâpatihicik / Their Example Showed Me the Way: A Cree Woman's Life Shaped by Two Cultures*. F. Ahenakew and H.C. Wolfart (Eds.). Edmonton: University of Alberta Press.

Mühlbauer, J. (2007). Word Order and the Interpretation of Nominals in Plains Cree. Ms, University of British Columbia.

Ogg, A. C. 1991. Connective particles and temporal cohesion in Plains Cree narrative. MA thesis, University of Manitoba.

Okimâsis, J. (2004). *Cree, Language of the Plains*. Regina, SK: University of Regina Press.

Okimâsis, J. & Wolvengrey, A. (2008). *How to spell it in Cree: The Standard Roman Orthography*. Regina, SK: *miywāsin* ink.

Proulx, P. (1988). The demonstrative pronouns of Proto-Algonquian. *International journal of American linguistics*, *54*(3), 309-330.

Rhodes, R., & Todd, E. (1981). Subarctic Algonquian languages. In *Subarctic* (Vol. 6, pp. 52-66). Washington: Smithsonian Institution.

Schmirler, Katherine & Harrigan, Atticus G. (2016). Word class frequencies according to corpora. Paper presented at the 3rd Prairie Workshop on Language and Linguistics in Regina, Saskatchewan, March 5, 2016.

Schmirler, K., Harrigan, A. G., Arppe, A., & Wolvengrey, A. (2017). A morphologically tagged corpus for Plains Cree. Paper presented at the 4th Annual Prairie Workshop on Language and Linguistics (PWoLL) in Saskatoon, Saskatchewan, March 18, 2017.

Snoek, C., Thunder, D., Lõo, K., Arppe, A., Lachler, J., Moshagen, S., & Trosterud, T. (2014). Modeling the Noun Morphology of Plains Cree. *ComputEL: Workshop on the use of computational methods in the study of endangered languages*, 52nd Annual Meeting of the ACL, Baltimore, Maryland, June 26, 2014.

Statistics Canada. (2015). *2011 Census of population*. Statistics Canada Catalogue no. 98-314-XCB20111047.

Trosterud, T. (2004). Porting morphological analysis and disambiguation to new languages. In *SALTMIL Workshop at LREC 2004: First Steps in Language Documentation for Minority Languages*, pp. 90-2.

Trosterud, T. (2005). Grammar-based language technology for the Sámi languages. *LULCL 2005*, 133.

Trosterud, T. (2007). Language technology for endangered languages: Sámi as a case study. (ms.)

Trosterud, T. (2009). A constraint grammar for Faroese. *NEALT PROCEEDINGS SERIES VOL. 8*, 1.

Universal Dependencies. (2017). Universal Dependencies v2. Retrieved from http://universaldependencies.org/.

University of Helsinki. (2017). Connexorin suomen kielen dependenssijäsennin / Connexor Finnish language dependency parser. Retrieved from http://www.ling.helsinki.fi/atk/sovellusohj/parsers/fi-fdg.shtml, June 17, 2017.

Vandall, P. & Douquette, J. (1987). *wâskahikaniwiyiniw-âcimowina / Stories of the House People, Told by Peter Vandall and Joe Douquette*. F. Ahenakew (Ed.). Winnipeg: University of Manitoba Press.

Wagner, I., Cowell, A., & Hwang J. D. (2016). Applying Universal Dependency to the Arapaho language. Proceedings of LAW X, pp. 171-9.

Whitecalf, S. (1993). *kinêhiyawiwininiwaw nêhiyawêwin / The Cree Language is Our Identity: The La Ronge Lectures of Sarah Whitecalf*. H.C. Wolfart and F. Ahenakew (Eds.). Winnipeg: University of Manitoba Press.

Wolfart, H. C. (1973). *Plains Cree: A grammatical study* (Vol. 63.5). Philadelphia: American Philosophical Society.

Wolfart, H. C. (1996). Sketch of Cree, an Algonquian language. In *Handbook of American Indians. Volume 17: Languages* (Vol. 17, pp. 390-439). Washington: Smithsonian Institute.

Wolvengrey, A. (2001). *Cree: Words*. Regina: University of Regina Press.

Wolvengrey, A. E. (2011). *Semantic and pragmatic functions in Plains Cree syntax* (Unpublished doctoral dissertation). LOT. Retrieved from http://dare.uva.nl/record/1/342704.

Appendix A: Lexicosemantic coding to improve the Constraint Grammar for Plains Cree

Many syntactic descriptions of Plains Cree repeatedly hit upon the same point: lexical semantics are an important factor in the construction and understanding of a Plains Cree sentence. Coding semantic features of individual lexical items or sets of lexical items, whether within the morphological model or within the parser itself, would benefit the capabilities of a Constraint Grammar parser considerably. While semantic relationships, especially between verbs and nouns, can be quite complex (cf. Dowty, 1991), some general patterns can improve the identification of relationships.

Prepositional phrases are a simple example of the benefits of semantic information. Like demonstratives and nouns, there are also syntactic relationships between prepositions, or locative particles, and locative nouns where linear orders are more fixed. Locative particles are a subset of particles that indicate spatial relationships and, when used in conjunction with a locative noun, always occur immediately before that noun. However, they may also occur without a noun just to indicate some spatial relationship. Some examples are given in (30).

(30) Prepositional phrases (Okimâsis, 2004, pp. 26-7)

a. *cîki sâkahikanihk*

cîki	sâkahikan	-ihk
near	lake.IN	LOC

‘near the lake’

b. *ispimihk mistikohk*

ispimihk	mistikw-	-ihk
up(stairs)	tree.AN	LOC

‘up (in) the tree’

However, as particles are not yet classified by function for Plains Cree,[40] these relationships may not be so straightforward to identify for the CG. Still, there are ways one might begin to identify these prepositions. First, many of them also contain the locative -*ihk*; the database used for our analyser, which underlies Wolvengrey’s (2001) dictionary of Cree, indicates that there are 83 particles containing -*ihk*. This small set could easily be manually coded for spatial functions as per their glosses. There are other locative particles such as *cîki* ‘near’ or *sisonê* ‘along’ that do not contain -*ihk*, but Cree textbooks such as Okimâsis (2004) and Ratt (2016) contain substantial lists of these preposition-like particles, which could also be extracted for coding as spatial/prepositional in the database. These tasks may not lead to an exhaustive list of Plains Cree prepositional particles, but would still improve the coverage of the CG.

Alongside the prepositions, there are also several postpositional locative particles in Cree: *isi* ‘toward, thus, thither’ and *ohci* ‘from, by means of, thence’ are common, while others such as *isko* are variable and may occur before or after other material.[41] These can also occur with

[40] Classification of all Plains Cree particles, e.g. spatial, prepositional, temporal, may prove useful for a CG analysis, though only prepositional particles seem to demonstrate such strict linear relationships.

[41] The postpositional status of *isi* and *ohci* is not unsurprising; the stems of these particles are *it*- and *oht*- respectively, which are members of the class of relative roots in Cree, which require some sort of antecedent. As such, the locative nouns occur before the postpositions; however, *isi* and *ohci* need not specify only spatial

locative nouns, though in these cases the nouns occur before the postpositional particles, as in (31). However, unlike the prepositional particles, *isi*, *ohci*, and *isko* do not necessarily carry spatial meaning on their own, but can be used with other, non-locative words as well. For example, *isi* can be used to mean ‘thus’ or ‘in such a way’, *ohci* can be used as ‘by means of’, and *isko* can be used as ‘until’ or ‘up to’ in terms of space, time, quantity, etc. Examples are given in (32).

(31) Postpositional phrases

a. *ôtênâhk ohci*

ôtênaw	-ihk	ohci
town.IN	LOC	from

‘from town’

b. *kiskinwahamâtowikamikohk isi*

kiskinwahamâtowikamikw-	-ihk	isi
school.IN	LOC	toward

‘to school’

(32) Non-spatial uses of *isi*, *ohci*, and *isko*

a. *môy ôhci-môhcowiw oskinîkiskwêw… ta-m[ê]tawâkâtikot nânitaw isi nâpêwa*

môy	ôhci-	môhcowi-	-w	oskinîkiskwêw	…
NEG	PST	be.foolish.VAI	3SG	young.woman.NA	
ta-	mêtawâkât-		-ikot	nânitaw isi	
CNJ	make.a.mockery.VTA		3′>3SG	in.any.way	
nâpêw		-a			
man.NA		OBV			

‘young women did not play around… so that men would make fools of them in any way’ (adapted from Ahenakew, 2000, pp. 44-5)

b. *maskihkiy êkota t-ôsîhtâyan niyaw ohci*

maskihkiy	êkota	ta-	osîhtâ-	-yan	n-	iyaw
medicine.NI	there	CNJ	make.VTI	2SG	1SG	body.NI
ohci						
by.means.of						

‘for you to make medicine there from my body’ (adapted from Ahenakew, 2000, pp. 64-5)

c. *ê-niyânano-kîsikâk isko mân ê-atoskêt*

ê-	niyânano-kîsikâ-	-k	isko	mâna	ê-	atoskê-	-t
CNJ	Friday.VII	3SG	until	usually	CNJ	work.VAI	3SG

‘he usually worked through to Friday’ (adapted from Bear et al., 1998, pp. 260-1)

relationships and so locative marking is not required. Similarly, these roots occur in preverbs and verb stems, wherein the antecedent can be a particle, a noun, a clause, or quoted speech. Semantic coding for these (e.g. *ohci* ‘by means of’ indicating that an oblique noun may function as a tool by which an action is performed, *itwêw* ‘s/he says’ indicating that quotation marks within certain boundaries indicate quoted speech), may also improve the CG, though this is not feasible at this time.

Interrogative constructions also present more regular word-ordering principles in Plains Cree, though the semantic coding of question words, which are currently under the umbrella of particles, would be beneficial to the parsing of these sentences. The interrogative particle *cî*, which is used only to form polar interrogatives, always occurs in the second position. In Functional Discourse Grammar, this is identified as P^2 (and not, for example, P^{I+1}), because of the importance a second position appears to play in a number of languages (Wolvengrey, 2011, p. 304).[42] The word that occurs before *cî* is in focus, or that which is being questioned. In the examples given in (33), the element before *cî* could be stressed in the English gloss to indicate focus.

(33) Polar interrogatives (Wolvengrey, 2011, pp. 304-5)

a. *nôhtêhkatêw cî?*

nôhtêhkatê-	-w	cî
be.hungry.VAI	3SG	INTER

'Is s/he hungry?'

b. *otâkosîhk cî kî-takosin?*

otâkosîhk	cî	kî-	takosin	-Ø
yesterday	INTER	PST	arrive.VAI	3SG

'Did s/he arrive *yesterday*?'

c. *atâwêwikamikohk cî kiwi-itohtân?*

atâwêwikamikw-	-ihk	cî	ki-	wî-	itohtê-	-n
store.IN	LOC	INTER	2	FUT	go.VAI	1/2SG

'Are you going to *the store*?'

Though *cî* is used only for polar interrogatives, content interrogatives also have some set word orders as well. The question word (often formed in Cree with the element *tân-*) must occur in the initial, focused, position (Wolvengrey, 2011, p. 311). Some examples are given in (34). Note that unlike the polar interrogatives above, all content interrogatives occur with conjunct verbs, though either *ê-* or *kâ-* may be used (Cook, 2014, pp. 235-8; Blain, 1997, p. 69).

(34) Content interrogatives (Wolvengrey 2011, pp. 311-3)

a. *awîna ê-kî-pakamahwat?*

awîna	ê-	kî	pakamahw-	-at
who.SG	CNJ	PST	hit.VTA	2SG>3SG

'Who did you hit?'

b. *kîkwây ê-kî-pakamahaman?*

kîkwây	ê-	kî-	pakamah-	-am	-an
what.SG	CNJ	PST	hit.VTI	THM	2SG

'What did you hit?'

[42] In Plains Cree, P^2 is also relevant to some coordinating particles (e.g. *êkwa* 'and', *mâka* 'but' [though both words have numerous functions]) and to emphatic particles such as *oti* and *ani*, both of which emphasise the preceding (topical or focused) element (Wolvengrey, 2011, p. 290).

c. *tânispî ê-wî-sipwêhtêyan?*
tânispî ê- wî- sipwêhtê- -yan
when CNJ FUT leave.VAI 2SG
'When are you going to leave?'

d. *tânitahto (masinahikana) ê-kî-atâwêyan?*
tânitahto (masinahikan -a) ê- kî- atâwê- -yan
how.many (book.IN PL) CNJ PST buy.VAI 2SG
'How many (books) did you buy?'

For both content and polar interrogatives, the CG could also make us of the clause boundary marker <?>, as the corpus makes use of standard English punctuation for the Cree texts. This can also allow for a distinction between interrogatives like those above and embedded *wh*-clauses. In the investigation of clause structure in Plains Cree, the punctuation in the corpus gives clues not only about sentences, but about clause boundaries (e.g. commas) within sentences as well. Intonational pauses, for example, can be used with respect to word order in Plains Cree (e.g. Mühlbauer, 2007); this aligns with the clausal template given in Wolvengrey (2011) in the Functional Discourse Grammar framework, which separates pre- and post-clausal segments with commas. In the CG, we can take such information to be peripheral to the clause, as pre-clausal information orients the clause, and post-clausal information contains afterthoughts or clarification, but neither need to be considered arguments within the clause proper.

At all levels of Plains Cree syntax, coordinators can also occur. The most common of these are *êkwa* 'and, then', *mâka* 'but', and *âhpô* 'or'. These can occur at the sentence level; *êkwa* is often used to link two sections of discourse together (e.g. "and then…") (e.g. Ogg, 1991). They can link clauses, nouns, noun phrases, verb phrases, etc. The identification of which type of unit a coordinator is linking may prove to be an interesting challenge for the CG, especially as *êkwa* and *mâka* are also commonly displaced to P^2 (the second position) of a given phrase (Wolvengrey, 2011). As for prepositional and question particles, all coordinating particles are not yet marked as such and so cannot be readily identified by the parser.

For the identification of arguments, one can rely primarily on the verb class and inflections to find relevant lexical items in the clause. While the verb class information is sufficient to determine which types of arguments are expected (animate, inanimate, actors, and goals), semantic information for subclasses may be useful for identifying ditransitive verbs, such as *asamêw* 's/he feeds s.t. to s.o.', exemplified in §2.2.1. For instance, many ditransitive verbs are benefactives that take both a recipient (that is, as the goal) plus an indirect object or theme, which may be either animate or inanimate. As benefactives are often derived with the suffix *-amaw*, these may easily be found in the lexicon coded as such for the reference to argument structure for the CG (Wolvengrey, 2011, p. 43; Wolfart, 1973, p. 61). The VAI *minihkwêw*, 's/he drinks', mentioned in §5.2., could also be coded for an expected oblique, something that is drunk—this verb has the connotation of 'drinking alcohol' (not unlike English); without an overt object somewhere in the discourse, alcohol is assumed and so many speakers specify the object as though it were a VTI (Wolvengrey, 2001). As the parsed corpus for Plains Cree is examined manually, various other verbs that seem to occur with specific thematic roles may also be noted and coded appropriately.

The current iteration of the CG also does not account for relationships between verbs, nouns, and embedded clauses. Independent and conjunct clauses offer some distinction here:

independent clauses can only be matrix clauses, and therefore cannot occur as relative clauses or clausal complements. Wolfart (1973) groups independent and imperative verbs together in contrast to conjunct and future conditional; the former can form clauses that can stand on their own while the latter require embedding.[43] In the most straightforward of sentences, wherein there is one independent and one conjunct clause, it is easy to indicate which is the matrix clause (independent), and which is embedded (conjunct). However, the distinction between independent and conjunct is far more complex than matrix vs. embedded clauses: conjunct clauses frequently occur as matrix clauses, not only as embedded clauses.

Cook (2014) conducted an in-depth investigation of independent and conjunct clauses, referring to them instead as indexical and anaphoric clauses respectively. Indexical clauses are evaluated against a speech situation and anaphoric clauses against a given context, which may be linguistic (e.g. a matrix clause) or extralinguistic. This distinction makes it possible for conjunct clauses to occur as matrix clauses: they are "embedded" in a real-world context. For example, conjunct verbs are used in the content interrogatives exemplified above: the fact that someone or something was hit, or that books were bought, can be considered known information, and instead *who*, *what*, or *how many* is what is being questioned. Unfortunately for a CG, this distinction does not necessarily simplify the situation or allow for easier recognition of interclausal dependencies. However, different conjunct preverbs, which are marked in the morphological model, can be used to determine whether a verb is embedded. While the *ê-* preverb can occur as a matrix clause, most other instances of conjunct verbs must be syntactically embedded. For instance, *kâ-* (discussed in more detail below), must be embedded, as must any conjunct verb containing initial change[44]—but still, this embedding need not be linguistic. Conjunct verbs may also occur with other conjunct prefixes, or no prefix at all, and these must also be embedded. Similarly, future conditional verbs are always in embedded clauses (Cook, 2014, p. 26-8). As such, matrix vs. embedded clause distinctions can be reliably determined from linguistic context only when both an independent and conjunct verb occur, though the preverb *kâ-* may offer a clue for relative clauses.

Wolfart (1973, p. 46) describes four functions for conjunct clauses: narrative (e.g. a temporal setting, not agreeing with an argument), participial (e.g. a relative clause describing something about an argument of a main verb), substantive (e.g. an adjunct to a verb), and focal (e.g. the conjunct verb occurs with a particle that focuses some aspect of the action). For narrative conjuncts, these would easily be identifiable if temporal VIIs were coded as such semantically; when all VIIs are indistinguishable, as is true for the current CG parser, temporal and other impersonal verbs (e.g. weather verbs) may incorrectly be identified as the governors of nominal or pronominal arguments. Relative clauses may also be identified in two main ways. First, the conjunct preverb *kâ-* is used in many relative clauses; however, this is not the only function of *kâ-*, and the more neutral conjunct preverb *ê-* may also be used for relative clauses, so this can only be a tendency, not a categorical rule. Second, if a verb in the clause not containing *kâ-* was evaluated first, any existing relationships between nouns and the main verb could be identified; if a *kâ-* conjunct verb matched in person and number features to a noun or

---

[43] Independent and conjunct clauses have also been contrasted in terms of simple vs. progressive aspect. Okimâsis (2004), for example, has combined embedding and aspect to exemplify the differences in English glosses: e.g., *nimâton* 'I cry' vs. *ê-mâtoyân* 'as I am crying'.

[44] Initial change is a change to the first vowel of a conjunct or future conditional verb (e.g. *i* > *ê*, *â* > *iya*) (Wolfart, 1973, 82-3). It is not yet implemented in the morphological analyser, as it is a rarer and rather archaic feature of Plains Cree, but can be referenced by the CG as it has been included in the manually-verified morphological Gold Standard.

pronoun, it would likely be a relative clause describing that element, though the element may also be an argument of another verb.

Further means of recognizing embedded clauses would require semantic classification of subordinating particles, though still these would not occur in every instance of an embedded clause. Any subordinating particle can only be used with conjunct verbs; these include *osâm* 'reason, because', *iyikohk* 'as far as', *kiyâm* 'although', *pâmwayês* 'before', *mayaw* 'as soon as', *âta* 'although', and *ayis* 'for, because', among others (Cook, 2014, pp. 60-1).

For clausal complements, individual verbs, based on their semantics, may be more likely to take clausal complements. Verbs that refer to states of mind (*think, like, know*) and speaking (*tell*) are likely candidates for taking clausal complements. For example, *kiskêyihtam* 's/he knows something' can occur with no overt arguments as in (35)a., a demonstrative as in (35)b., or with a clausal complement, as in (35)c.[45] As (35)c. shows, the demonstrative may also occur with a complement clause.

(35) 'to know' (Cook, 2014, pp. 209-10)

a. *nikiskêyihtên.*

ni-	kiskêyiht-	-ê	-n
1	know.VTI	THM	1/2SG

'I know it.'

b. *nikiskêyihtên ôma.*

ni-	kiskêyiht-	-ê	-n	ôma
1	know.VTI	THM	1/2SG	this.IN

'I know this.'

c. *nikiskêyihtên (ôma) ê-wî-kîwêyan.*

ni-	kiskêyiht-	-ê	-n	(ôma)	ê-	wî-	kîwê-
1	know.VTI	THM	1/2SG	(this.IN)	CNJ	FUT	go.home.VAI
-yan							
2SG							

'I know that you're going to go home.'

Semantic features such as those described above will allow for considerable improvement to a Constraint Grammar parser for Plains Cree. Some of these features may be best coded in the lexicon of the morphological analyser, such as certain subtypes of pronouns or particles (e.g. interrogatives), while others may be best indicated in a set within the parser (e.g. benefactives, prepositional particles, etc.). As more features become available for reference in the CG parser, its capabilities will improve to the point that full dependency analyses may be automatically produced for each sentence for inclusion within a corpus. Full dependency analyses can also be used for further large-scale investigations of Plains Cree syntax.

---

[45] The VTA counterparts to such verbs, e.g. *kiskêyimêw* 's/he knows s.o.' can also be used with clausal complements with animate actors; the actor of the embedded clause becomes the goal of the matrix VTA (Dahlstrom, 1991, p. 76; Wolvengrey, 2011, p. 186; Cook, 2014, p. 206).

Appendix B: Further phrase order frequency counts

While only the most frequent phrase order results were presented in §5.2, many more orders present themselves in the corpus: there are multiple verbs within clauses and multiple actors and goals. Many of these orders occur only once, and as such may represent ambiguity in the Gold Standard that introduced errors when verbs were tagged in the parser. Furthermore, phrases containing obliques and prepositions and those without verbs (e.g. just a noun phrase or prepositional phrase) were excluded above, as the focus was verbs and their arguments. These are retained here, though it must be noted that the development of the current CG parser did not focus on refining the constraints that apply to obliques and prepositions and so these counts likely do not accurately reflect their occurrence in the corpus. As there are over 1,000 different phrase orders attested in the corpus, only orders that occur four or more times are presented in Table 10.

Table 10 Phrase orders across the corpus

Frequency	Phrase order			% of 19,734
4865	@PRED-AI			24.7
2128	@PRED-TA			10.8
1432	@PRED-TI			7.3
801	@PRED-II			4.1
539	@PRED-TA	@<GOAL		2.7
521	@ACTOR>	@PRED-AI		2.6
503	@PRED-AI	@<ACTOR		2.5
478	@N>			2.4
450	@GOAL>	@PRED-TI		2.3
386	@GOAL>	@PRED-TA		2.0
324	@PRED-TI	@<GOAL		1.6
242	@PRED-AI	@PRED-AI		1.2
211	@ACTOR>	@PRED-II		1.1
184	@<N			0.9
177	@OBL>	@PRED-AI		0.9
173	@ACTOR>	@PRED-TA		0.9
162	@PRED-AI	@N>	@<ACTOR	0.8
159	@PRED-II	@<ACTOR		0.8
148	@PRED-TA	@<ACTOR		0.7
132	@PRED-TA	@N>	@<GOAL	0.7
122	@PRED-AI	@<OBL		0.6
116	@PRED-TA	@PRED-AI		0.6
115	@OBL>	@PRED-TA		0.6
114	@PRED-TI	@PRED-AI		0.6
111	@ACTOR>	@PRED-TI		0.6
103	@PRED-TI	@<ACTOR		0.5
79	@PRED-TA	@<OBL		0.4
76	@PRED-TI	@N>	@<GOAL	0.4
72	@PRED-II	@PRED-AI		0.4
71	@PRED-TA	@PRED-TA		0.4
64	@PRED-AI	@PRED-TA		0.3
64	@ACTOR>	@ACTOR>	@PRED-AI	0.3
59	@PRED-TI	@PRED-TI		0.3
57	@GOAL>	@PRED-AI		0.3
54	@PRED-TA	@N>	@<ACTOR	0.3
54	@<P			0.3
52	@ACTOR>	@PRED-AI	@<ACTOR	0.3

45	@PRED-TI	@PRED-TA		0.2
39	@PRED-AI	@PRED-TI		0.2
39	@N>	@ACTOR>	@PRED-AI	0.2
38	@PRED-AI	@<ACTOR	@<ACTOR	0.2
35	@PRED-TI	@PRED-II		0.2
35	@PRED-II	@N>	@<ACTOR	0.2
35	@N>	@PRED-AI		0.2
35	@ACTOR>	@PRED-TA	@<GOAL	0.2
35	@ACTOR>	@<N	@PRED-AI	0.2
34	@PRED-AI	@N>	@<OBL	0.2
33	@PRED-AI	@PRED-II		0.2
33	@ACTOR>	@PRED-TI	@<GOAL	0.2
32	@<P	@PRED-AI		0.2
32	@N>	@GOAL>	@PRED-TI	0.2
31	@N>	@<N		0.2
30	@GOAL>	@PRED-TI	@<GOAL	0.2
28	@GOAL>	@PRED-TA	@<GOAL	0.1
27	@GOAL>	@<N	@PRED-TI	0.1
26	@PRED-TA	@PRED-TI		0.1
26	@GOAL>	@GOAL>	@PRED-TI	0.1
25	@P>			0.1
24	@PRED-TA	@<GOAL	@PRED-AI	0.1
24	@PRED-II	@PRED-TI		0.1
23	@PRED-AI	@<ACTOR	@PRED-AI	0.1
23	@N>	@GOAL>	@PRED-TA	0.1
22	@PRED-TI	@N>	@<ACTOR	0.1
22	@PRED-TI	@<GOAL	@PRED-AI	0.1
22	@PRED-TA	@<GOAL	@<GOAL	0.1
21	@PRED-II	@PRED-TA		0.1
21	@PRED-II	@PRED-II		0.1
21	@PRED-AI	@PRED-TA	@<GOAL	0.1
21	@PRED-AI	@GOAL>	@PRED-TI	0.1
21	@ACTOR>	@PRED-TA	@<ACTOR	0.1
20	@GOAL>	@GOAL>	@PRED-TA	0.1
20	@ACTOR>	@PRED-AI	@PRED-AI	0.1
20	@ACTOR>	@GOAL>	@PRED-TI	0.1
19	@PRED-TI	@<GOAL	@PRED-TI	0.1
19	@PRED-II	@<ACTOR	@PRED-AI	0.1
19	@P>	@PRED-AI		0.1
19	@GOAL>	@PRED-TI	@PRED-TI	0.1
18	@PRED-AI	@GOAL>	@PRED-TA	0.09
17	@PRED-TA	@<GOAL	@PRED-TA	0.09
17	@PRED-AI	@PRED-AI	@<ACTOR	0.09
17	@OBL>	@PRED-TA	@<GOAL	0.09
17	@N>	@ACTOR>	@PRED-TA	0.09
17	@ACTOR>	@GOAL>	@PRED-TA	0.09
16	@N>	@N>		0.08
16	@GOAL>	@PRED-TI	@PRED-AI	0.08
16	@ACTOR>	@<N	@PRED-II	0.08
15	@GOAL>	@PRED-TA	@PRED-AI	0.08
14	@PRED-TI	@<ACTOR	@<GOAL	0.07
14	@PRED-TA	@N>	@<OBL	0.07
14	@PRED-TA	@ACTOR>	@PRED-AI	0.07
14	@N>	@PRED-TA		0.07
14	@ACTOR>	@PRED-II	@<ACTOR	0.07
13	@PRED-TI	@<GOAL	@<GOAL	0.07

13	@PRED-TI	@<ACTOR	@PRED-AI		0.07
13	@PRED-II	@PRED-TA	@<GOAL		0.07
13	@PRED-AI	@PRED-AI	@PRED-AI		0.07
13	@<P	@PRED-TA			0.07
13	@N>	@ACTOR>	@PRED-II		0.07
13	@GOAL>	@<N	@PRED-TA		0.07
13	@ACTOR>	@PRED-AI	@N>	@<ACTOR	0.07
12	@PRED-TI	@<OBL			0.06
12	@PRED-TA	@N>			0.06
12	@PRED-TA	@<ACTOR	@<GOAL		0.06
12	@PRED-AI	@<P			0.06
12	@GOAL>	@PRED-TI	@<ACTOR		0.06
12	@GOAL>	@PRED-TA	@<ACTOR		0.06
12	@ACTOR>	@PRED-AI	@<OBL		0.06
11	@PRED-TI	@<GOAL	@PRED-TA		0.06
11	@PRED-AI	@<GOAL			0.06
11	@N>	@PRED-TI			0.06
11	@GOAL>	@PRED-TA	@N>	@<GOAL	0.06
11	@GOAL>	@ACTOR>	@PRED-TA		0.06
10	@PRED-TI	@PRED-AI	@<ACTOR		0.05
10	@PRED-TA	@<GOAL	@<OBL		0.05
10	@PRED-II	@<ACTOR	@PRED-TA		0.05
10	@PRED-AI	@PRED-TI	@<GOAL		0.05
10	@PRED-AI	@<ACTOR	@<OBL		0.05
10	@OBL>	@OBL>	@PRED-AI		0.05
10	@N>	@OBL>	@PRED-TA		0.05
10	@GOAL>	@ACTOR>	@PRED-TI		0.05
10	@ACTOR>	@PRED-II	@PRED-TA		0.05
10	@ACTOR>	@PRED-II	@PRED-AI		0.05
10	@ACTOR>	@ACTOR>	@PRED-TI		0.05
9	@PRED-TI	@PRED-TI	@<GOAL		0.05
9	@PRED-TI	@ACTOR>	@PRED-AI		0.05
9	@PRED-TA	@PRED-II			0.05
9	@PRED-TA	@<ACTOR	@<ACTOR		0.05
9	@PRED-II	@PRED-AI	@<ACTOR		0.05
9	@PRED-AI	@ACTOR>	@PRED-II		0.05
9	@OBL>	@PRED-AI	@<ACTOR		0.05
9	@N>	@PRED-II			0.05
9	@GOAL>	@GOAL>	@PRED-AI		0.05
9	@ACTOR>	@ACTOR>	@PRED-TA		0.05
8	@PRED-TI	@N>	@<OBL		0.04
8	@PRED-TA	@PRED-TI	@<GOAL		0.04
8	@PRED-TA	@GOAL>	@PRED-TA		0.04
8	@PRED-TA	@<ACTOR	@PRED-AI		0.04
8	@PRED-II	@N>			0.04
8	@PRED-II	@GOAL>	@PRED-TI		0.04
8	@PRED-AI	@N>	@<ACTOR	@<N	0.04
8	@N>	@ACTOR>	@PRED-TI		0.04
8	@GOAL>	@PRED-TA	@PRED-TA		0.04
8	@ACTOR>	@<N	@PRED-TA		0.04
8	@ACTOR>	@ACTOR>	@PRED-II		0.04
7	@PRED-TI	@PRED-TA	@<GOAL		0.04
7	@PRED-TI	@<GOAL	@<ACTOR		0.04
7	@PRED-TA	@PRED-TA	@<GOAL		0.04
7	@PRED-TA	@ACTOR>	@PRED-TA		0.04
7	@PRED-AI	@<ACTOR	@PRED-TA		0.04

7	@OBL>	@PRED-TA	@PRED-TA		0.04
7	@OBL>	@PRED-TA	@<ACTOR		0.04
7	@OBL>	@<N	@PRED-AI		0.04
7	@N>	@OBL>	@PRED-AI		0.04
7	@GOAL>	@PRED-TI	@PRED-II		0.04
7	@GOAL>	@OBL>	@PRED-TA		0.04
7	@ACTOR>	@PRED-TI	@PRED-AI		0.04
7	@ACTOR>	@PRED-TA	@PRED-AI		0.04
7	@ACTOR>	@PRED-AI	@PRED-TA		0.04
6	@PRED-TI	@GOAL>	@PRED-TA		0.03
6	@PRED-TI	@<GOAL	@<N		0.03
6	@PRED-TA	@PRED-AI	@PRED-AI		0.03
6	@PRED-TA	@<OBL	@<OBL		0.03
6	@PRED-TA	@<GOAL	@PRED-TA	@<GOAL	0.03
6	@PRED-AI	@P>			0.03
6	@PRED-AI	@<ACTOR	@PRED-TA	@<GOAL	0.03
6	@PRED-AI	@ACTOR>	@PRED-TA		0.03
6	@PRED-AI	@ACTOR>	@PRED-AI		0.03
6	@PRED-AI	@<ACTOR	@<N		0.03
6	@<P	@PRED-TI			0.03
6	@OBL>	@PRED-AI	@PRED-AI		0.03
6	@OBL>	@PRED-AI	@<OBL		0.03
6	@GOAL>	@PRED-TI	@PRED-TA		0.03
6	@ACTOR>	@PRED-TA	@<OBL		0.03
6	@ACTOR>	@PRED-TA	@N>	@<GOAL	0.03
6	@ACTOR>	@OBL>	@PRED-AI		0.03
5	@PRED-TI	@PRED-TA	@<ACTOR		0.03
5	@PRED-TI	@PRED-II	@<ACTOR		0.03
5	@PRED-TI	@ACTOR>	@PRED-TA		0.03
5	@PRED-TI	@ACTOR>	@PRED-II		0.03
5	@PRED-TA	@PRED-TA	@N>	@<GOAL	0.03
5	@PRED-TA	@PRED-TA	@<ACTOR		0.03
5	@PRED-TA	@PRED-AI	@<ACTOR		0.03
5	@PRED-TA	@<P			0.03
5	@PRED-TA	@<OBL	@PRED-TA		0.03
5	@PRED-TA	@<GOAL	@PRED-TI		0.03
5	@PRED-TA	@<GOAL	@<N		0.03
5	@PRED-TA	@<GOAL	@<ACTOR		0.03
5	@PRED-II	@PRED-AI	@PRED-AI		0.03
5	@PRED-AI	@PRED-II	@PRED-AI		0.03
5	@PRED-AI	@PRED-AI	@<OBL		0.03
5	@PRED-AI	@<OBL	@<N		0.03
5	@PRED-AI	@N>			0.03
5	@P>	@PRED-TI			0.03
5	@OBL>	@PRED-TI			0.03
5	@OBL>	@PRED-TA	@<OBL		0.03
5	@OBL>	@PRED-AI	@PRED-II		0.03
5	@OBL>	@<N	@PRED-TA		0.03
5	@OBL>	@GOAL>	@PRED-TA		0.03
5	@OBL>	@ACTOR>	@PRED-AI		0.03
5	@N>	@ACTOR>	@<N	@PRED-AI	0.03
5	@GOAL>	@PRED-TI	@N>	@<GOAL	0.03
5	@GOAL>	@GOAL>	@PRED-TA	@<GOAL	0.03
5	@ACTOR>	@PRED-TI	@PRED-TI		0.03
5	@ACTOR>	@PRED-TI	@PRED-TA		0.03
5	@ACTOR>	@PRED-TI	@N>	@<GOAL	0.03

5	@ACTOR>	@PRED-TI	@<ACTOR		0.03
5	@ACTOR>	@ACTOR>	@PRED-AI	@<ACTOR	0.03
4	@PRED-TA	@PRED-AI	@<OBL		0.02
4	@PRED-TA	@P>			0.02
4	@PRED-TA	@N>	@<GOAL	@PRED-AI	0.02
4	@PRED-TA	@GOAL>	@PRED-TI		0.02
4	@PRED-TA	@<GOAL	@<GOAL	@PRED-AI	0.02
4	@PRED-TA	@<ACTOR	@<N		0.02
4	@PRED-II	@PRED-TI	@<GOAL		0.02
4	@PRED-II	@<ACTOR	@PRED-TI		0.02
4	@PRED-AI	@PRED-II	@<ACTOR		0.02
4	@PRED-AI	@PRED-AI	@N>	@<ACTOR	0.02
4	@PRED-AI	@<P	@PRED-AI		0.02
4	@PRED-AI	@<OBL	@PRED-AI		0.02
4	@PRED-AI	@<OBL	@<OBL		0.02
4	@PRED-AI	@<ACTOR	@<ACTOR	@PRED-AI	0.02
4	@OBL>	@PRED-TA	@N>	@<ACTOR	0.02
4	@N>	@PRED-AI	@<ACTOR		0.02
4	@GOAL>	@PRED-TI	@PRED-AI	@<ACTOR	0.02
4	@GOAL>	@PRED-TA	@N>	@<ACTOR	0.02
4	@GOAL>	@PRED-AI	@<GOAL		0.02
4	@GOAL>	@GOAL>	@GOAL>	@PRED-TA	0.02
4	@ACTOR>	@PRED-AI	@PRED-AI	@<ACTOR	0.02
4	@ACTOR>	@<N	@PRED-TI	@<GOAL	0.02
4	@ACTOR>	@N>	@ACTOR>	@PRED-AI	0.02
4	@ACTOR>	@GOAL>	@PRED-AI		0.02
4	@ACTOR>	@ACTOR>	@PRED-TI	@<GOAL	0.02
4	@ACTOR>	@ACTOR>	@PRED-AI	@PRED-AI	0.02

Table 11 includes VTA phrase orders that occur at least twice in the corpus. As in §5.2.3, phrases containing other verb types are not included. Phrases that include both a goal and an oblique (that is, likely a direct and indirect object) are presented in boldface.

Table 11 VTA phrase orders

Frequency	Phrase order			% of 4,551
2128	@PRED-TA			46.8
539	@PRED-TA	@<GOAL		11.8
386	@GOAL>	@PRED-TA		8.5
173	@ACTOR>	@PRED-TA		3.8
148	@PRED-TA	@<ACTOR		3.3
132	@PRED-TA	@N>	@<GOAL	2.9
115	@OBL>	@PRED-TA		2.5
79	@PRED-TA	@<OBL		1.7
71	@PRED-TA	@PRED-TA		1.6
54	@PRED-TA	@N>	@<ACTOR	1.2
35	@ACTOR>	@PRED-TA	@<GOAL	0.8
28	@GOAL>	@PRED-TA	@<GOAL	0.6
23	@N>	@GOAL>	@PRED-TA	0.5
22	@PRED-TA	@<GOAL	@<GOAL	0.5
21	@ACTOR>	@PRED-TA	@<ACTOR	0.5
20	@GOAL>	@GOAL>	@PRED-TA	0.4
17	@PRED-TA	@<GOAL	@PRED-TA	0.4
17	**@OBL>**	**@PRED-TA**	**@<GOAL**	0.4
17	@N>	@ACTOR>	@PRED-TA	0.4

17	@ACTOR>	@GOAL>	@PRED-TA		0.4
14	@PRED-TA	@N>	@<OBL		0.3
14	@N>	@PRED-TA			0.3
13	@<P	@PRED-TA			0.3
13	@GOAL>	@<N	@PRED-TA		0.3
12	@PRED-TA	@N>			0.3
12	@PRED-TA	@<ACTOR	@<GOAL		0.3
12	@GOAL>	@PRED-TA	@<ACTOR		0.3
11	@GOAL>	@PRED-TA	@N>	@<GOAL	0.2
11	@GOAL>	@ACTOR>	@PRED-TA		0.2
10	**@PRED-TA**	**@<GOAL**	**@<OBL**		0.2
10	@N>	@OBL>	@PRED-TA		0.2
9	@PRED-TA	@<ACTOR	@<ACTOR		0.2
9	@ACTOR>	@ACTOR>	@PRED-TA		0.2
8	@PRED-TA	@GOAL>	@PRED-TA		0.2
8	@GOAL>	@PRED-TA	@PRED-TA		0.2
8	@ACTOR>	@<N	@PRED-TA		0.2
7	@PRED-TA	@PRED-TA	@<GOAL		0.2
7	@PRED-TA	@ACTOR>	@PRED-TA		0.2
7	@OBL>	@PRED-TA	@PRED-TA		0.2
7	@OBL>	@PRED-TA	@<ACTOR		0.2
7	**@GOAL>**	**@OBL>**	**@PRED-TA**		0.2
6	@PRED-TA	@<OBL	@<OBL		0.1
6	@PRED-TA	@<GOAL	@PRED-TA	@<GOAL	0.1
6	@ACTOR>	@PRED-TA	@<OBL		0.1
6	@ACTOR>	@PRED-TA	@N>	@<GOAL	0.1
5	@PRED-TA	@PRED-TA	@N>	@<GOAL	0.1
5	@PRED-TA	@PRED-TA	@<ACTOR		0.1
5	@PRED-TA	@<P			0.1
5	@PRED-TA	@<OBL	@PRED-TA		0.1
5	@PRED-TA	@<GOAL	@<N		0.1
5	@PRED-TA	@<GOAL	@<ACTOR		0.1
5	@OBL>	@PRED-TA	@<OBL		0.1
5	@OBL>	@<N	@PRED-TA		0.1
5	**@OBL>**	**@GOAL>**	**@PRED-TA**		0.1
5	@GOAL>	@GOAL>	@PRED-TA	@<GOAL	0.1
4	@PRED-TA	@P>			0.1
4	@PRED-TA	@<ACTOR	@<N		0.1
4	@OBL>	@PRED-TA	@N>	@<ACTOR	0.1
4	@GOAL>	@PRED-TA	@N>	@<ACTOR	0.1
4	@GOAL>	@GOAL>	@GOAL>	@PRED-TA	0.1
3	**@PRED-TA**	**@<OBL**	**@<GOAL**		0.1
3	@PRED-TA	@<ACTOR	@<OBL		0.1
3	@P>	@PRED-TA			0.1
3	@OBL>	@OBL>	@PRED-TA		0.1
3	@GOAL>	@PRED-TA	@<P		0.1
3	@GOAL>	@PRED-TA	@<GOAL	@PRED-TA	0.1
3	@ACTOR>	@PRED-TA	@PRED-TA	@<GOAL	0.1
3	@ACTOR>	@PRED-TA	@PRED-TA		0.1
3	@ACTOR>	@ACTOR>	@PRED-TA	@<GOAL	0.1
2	@PRED-TA	@PRED-TA	@<ACTOR	@<GOAL	0.04
2	@PRED-TA	@N>	@PRED-TA		0.04
2	@PRED-TA	@N>	@<GOAL	@PRED-TA	0.04
2	@PRED-TA	@N>	@<GOAL	@<N	0.04
2	@PRED-TA	@N>	@ACTOR>	@PRED-TA	0.04
2	@PRED-TA	@N>	@<ACTOR	@PRED-TA	0.04

2	@PRED-TA	@N>	@<ACTOR	@N>	@<GOAL	0.04
2	@PRED-TA	@<N				0.04
2	@PRED-TA	@<GOAL	@N>	@<GOAL		0.04
2	@PRED-TA	@<GOAL	@N>			0.04
2	@PRED-TA	@<GOAL	@<GOAL	@<N		0.04
2	@PRED-TA	@ACTOR>	@PRED-TA	@<GOAL		0.04
2	@PRED-TA	@<ACTOR	@PRED-TA			0.04
2	@P>	@GOAL>	@PRED-TA			0.04
2	**@OBL>**	**@PRED-TA**	**@PRED-TA**	**@<GOAL**		0.04
2	@OBL>	@ACTOR>	@PRED-TA	@<ACTOR		0.04
2	@OBL>	@ACTOR>	@PRED-TA			0.04
2	@N>	@OBL>	@PRED-TA	@<ACTOR		0.04
2	@N>	@GOAL>	@<N	@PRED-TA		0.04
2	@N>	@ACTOR>	@PRED-TA	@<GOAL		0.04
2	@N>	@ACTOR>	@PRED-TA	@<ACTOR		0.04
2	@GOAL>	@PRED-TA	@PRED-TA	@<ACTOR		0.04
2	@GOAL>	@PRED-TA	@N>			0.04
2	@GOAL>	@PRED-TA	@<GOAL	@<GOAL		0.04
2	@GOAL>	@<P	@PRED-TA			0.04
2	@GOAL>	@GOAL>	@ACTOR>	@PRED-TA		0.04
2	@ACTOR>	@PRED-TA	@<P			0.04
2	@ACTOR>	@PRED-TA	@<GOAL	@<GOAL		0.04
2	@ACTOR>	@OBL>	@PRED-TA			0.04
2	@ACTOR>	@<N	@PRED-TA	@<GOAL		0.04
2	@ACTOR>	@<N	@PRED-TA	@<ACTOR		0.04

Below the full frequency lists for non-local direct and indirect VTAs are presented. In Table 12, non-local direct phrases are presented, including obliques and excluding other verb classes. Finally, the full list of phrase orders for non-local inverse VTAs is given in Table **13**. Again, obliques are included but other verb classes are excluded.

Table 12 Non-local direct VTA phrase orders

Frequency	Phrase order					% of 985
303	@PRED-TA					30.8
189	@PRED-TA	@<GOAL				19.2
133	@GOAL>	@PRED-TA				13.5
47	@ACTOR>	@PRED-TA				4.8
46	@PRED-TA	@N>	@<GOAL			4.7
37	@PRED-TA	@<ACTOR				3.8
29	@ACTOR>	@PRED-TA	@<GOAL			2.9
15	@PRED-TA	@N>	@<ACTOR			1.5
15	@OBL>	@PRED-TA				1.5
10	@PRED-TA	@<OBL				1.0
10	@ACTOR>	@GOAL>	@PRED-TA			1.0
9	@OBL>	@PRED-TA	@<GOAL			0.9
8	@GOAL>	@PRED-TA	@<GOAL			0.8
8	@GOAL>	@PRED-TA	@<ACTOR			0.8
8	@GOAL>	@ACTOR>	@PRED-TA			0.8
8	@ACTOR>	@PRED-TA	@<ACTOR			0.8
7	@PRED-TA	@<ACTOR	@<GOAL			0.7
6	@PRED-TA	@<GOAL	@<GOAL			0.6
5	@ACTOR>	@ACTOR>	@PRED-TA			0.5
4	@PRED-TA	@N>	@<OBL			0.4
4	@PRED-TA	@<GOAL	@<N			0.4
4	@N>	@GOAL>	@PRED-TA			0.4
4	@ACTOR>	@PRED-TA	@N>	@<GOAL		0.4
3	@PRED-TA	@<GOAL	@<OBL			0.3
3	@PRED-TA	@<ACTOR	@<ACTOR			0.3
3	@ACTOR>	@ACTOR>	@PRED-TA	@<GOAL		0.3
2	@PRED-TA	@<P				0.2
2	@PRED-TA	@N>	@<ACTOR	@N>	@<GOAL	0.2
2	@PRED-TA	@<GOAL	@<ACTOR			0.2
2	@PRED-TA	@<ACTOR	@<N			0.2
2	@<P	@PRED-TA				0.2
2	@OBL>	@GOAL>	@PRED-TA			0.2
2	@N>	@PRED-TA				0.2
2	@N>	@OBL>	@PRED-TA			0.2
2	@N>	@ACTOR>	@PRED-TA	@<GOAL		0.2
2	@N>	@ACTOR>	@PRED-TA			0.2
2	@GOAL>	@PRED-TA	@N>	@<GOAL		0.2
2	@GOAL>	@PRED-TA	@N>	@<ACTOR		0.2

2	@GOAL>	@<N	@PRED-TA				0.2
2	@GOAL>	@GOAL>	@PRED-TA				0.2
2	@ACTOR>	@<N	@PRED-TA	@<GOAL			0.2
1	@PRED-TA	@<OBL	@<OBL				0.1
1	@PRED-TA	@<OBL	@<GOAL	@<GOAL			0.1
1	@PRED-TA	@N>	@<GOAL	@<ACTOR			0.1
1	@PRED-TA	@N>	@<ACTOR	@<GOAL			0.1
1	@PRED-TA	@N>					0.1
1	@PRED-TA	@<ACTOR	@<OBL				0.1
1	@PRED-TA	@<ACTOR	@N>	@<ACTOR			0.1
1	@PRED-TA	@<ACTOR	@<GOAL	@<OBL			0.1
1	@PRED-TA	@<ACTOR	@<GOAL	@<ACTOR			0.1
1	@PRED-TA	@<ACTOR	@<ACTOR	@<N	@N>		0.1
1	@P>	@PRED-TA	@<GOAL				0.1
1	@P>	@GOAL>	@PRED-TA				0.1
1	@OBL>	@PRED-TA	@N>	@<ACTOR			0.1
1	@OBL>	@PRED-TA	@<ACTOR				0.1
1	@OBL>	@OBL>	@PRED-TA	@<GOAL			0.1
1	@OBL>	@ACTOR>	@PRED-TA	@N>	@<ACTOR		0.1
1	@OBL>	@ACTOR>	@PRED-TA				0.1
1	@N>	@OBL>	@PRED-TA	@<ACTOR			0.1
1	@N>	@N>	@GOAL>	@GOAL>	@PRED-TA		0.1
1	@N>	@GOAL>	@PRED-TA	@<ACTOR	@<GOAL		0.1
1	@N>	@GOAL>	@<N	@N>	@PRED-TA		0.1
1	@N>	@ACTOR>	@PRED-TA	@<ACTOR	@<N		0.1
1	@N>	@ACTOR>	@GOAL>	@PRED-TA			0.1
1	@GOAL>	@PRED-TA	@<P				0.1
1	@GOAL>	@PRED-TA	@<OBL	@<N			0.1
1	@GOAL>	@PRED-TA	@<ACTOR	@<GOAL			0.1
1	@GOAL>	@PRED-TA	@<ACTOR	@<ACTOR			0.1
1	@ACTOR>	@PRED-TA	@<OBL	@<GOAL			0.1
1	@ACTOR>	@PRED-TA	@N>	@<ACTOR	@<N		0.1
1	@ACTOR>	@PRED-TA	@<GOAL	@<OBL			0.1
1	@ACTOR>	@OBL>	@PRED-TA	@<GOAL			0.1
1	@ACTOR>	@<N	@PRED-TA	@<ACTOR			0.1
1	@ACTOR>	@<N	@PRED-TA				0.1
1	@ACTOR>	@<N	@ACTOR>	@PRED-TA			0.1
1	@ACTOR>	@GOAL>	@PRED-TA	@N>	@<ACTOR		0.1
1	@ACTOR>	@GOAL>	@<N	@PRED-TA	@<ACTOR		0.1
1	@ACTOR>	@ACTOR>	@ACTOR>	@PRED-TA	@N>	@<GOAL	0.1

Table 13 Non-local inverse VTA phrase orders

Frequency	Phrase order						% of 217
87	@PRED-TA						40.1
21	@PRED-TA	@<ACTOR					9.7
15	@PRED-TA	@<GOAL					6.9
15	@OBL>	@PRED-TA					6.9
15	@ACTOR>	@PRED-TA					6.9
9	@PRED-TA	@<OBL					4.1
7	@GOAL>	@PRED-TA					3.2
4	@PRED-TA	@N>	@<ACTOR				1.8
2	@PRED-TA	@<OBL	@<GOAL				0.9
2	@N>	@OBL>	@PRED-TA				0.9
2	@GOAL>	@PRED-TA	@N>	@<ACTOR			0.9
2	@GOAL>	@PRED-TA	@<ACTOR				0.9
2	@ACTOR>	@PRED-TA	@<OBL				0.9
2	@ACTOR>	@PRED-TA	@<ACTOR				0.9
1	@PRED-TA	@<P					0.5
1	@PRED-TA	@<OBL	@<OBL				0.5
1	@PRED-TA	@N>	@<OBL				0.5
1	@PRED-TA	@N>	@<GOAL	@N>	@<OBL		0.5
1	@PRED-TA	@N>	@<GOAL				0.5
1	@PRED-TA	@<GOAL	@<OBL				0.5
1	@PRED-TA	@<GOAL	@<GOAL	@<GOAL			0.5
1	@PRED-TA	@<GOAL	@<ACTOR	@<ACTOR			0.5
1	@PRED-TA	@<ACTOR	@<OBL				0.5
1	@PRED-TA	@<ACTOR	@<GOAL				0.5
1	@PRED-TA	@<ACTOR	@<ACTOR				0.5
1	@P>	@PRED-TA					0.5
1	@OBL>	@PRED-TA	@<GOAL				0.5
1	@OBL>	@PRED-TA	@<ACTOR	@<N			0.5
1	@OBL>	@PRED-TA	@<ACTOR				0.5
1	@OBL>	@OBL>	@PRED-TA				0.5
1	@OBL>	@OBL>	@GOAL>	@PRED-TA			0.5
1	@OBL>	@GOAL>	@PRED-TA				0.5
1	@N>	@OBL>	@PRED-TA	@N>	@<GOAL	@<N	0.5
1	@N>	@GOAL>	@OBL>	@PRED-TA			0.5
1	@N>	@ACTOR>	@PRED-TA	@<ACTOR			0.5
1	@GOAL>	@PRED-TA	@<P				0.5
1	@GOAL>	@<N	@PRED-TA	@<ACTOR			0.5

1	@GOAL>	@GOAL>	@PRED-TA	@<ACTOR		0.5
1	@GOAL>	@GOAL>	@ACTOR>	@PRED-TA		0.5
1	@GOAL>	@ACTOR>	@PRED-TA			0.5
1	@GOAL>	@ACTOR>	@OBL>	@PRED-TA	@<OBL	0.5
1	@ACTOR>	@PRED-TA	@<P			0.5
1	@ACTOR>	@PRED-TA	@N>	@<ACTOR		0.5
1	@ACTOR>	@PRED-TA	@<ACTOR	@<ACTOR		0.5
1	@ACTOR>	@<N	@PRED-TA			0.5
1	@ACTOR>	@ACTOR>	@ACTOR>	@PRED-TA		0.5