Analyzing and generating English phrases with finite-state methods to match and translate inflected Plains Cree (*nêhiyawêwin*) word-forms

Antti Arppe
Alberta Language Technology Lab
Department of Linguistics, University of Alberta
arppe@ualberta.ca

Abstract

This paper presents two finite-state transducer tools, which can be used to analyze or generate simple English verb and noun phrases, that can be mapped with inflected Plains Cree (*nêhiyawêwin)* verb and noun forms. These tools support fetching an inflected Cree word-form directly with an appropriate plain English phrase, and conversely providing a rough translation of an inflected Cree word-form. Such functionalities can be used to improve the user friendliness of on-line dictionaries. The tools are extendable to other similarly morphologically complex languages.

1 Introduction

Exemplifying with the pairing of a morphologically complex Indigenous language spoken on the Western plains of Canada and the corresponding morphologically simpler majority language, namely Plains Cree (*nêhiyawêwin*; ISO: crk) and English, this paper presents computational tools using finite-state transducer (FST) technology for analyzing and generating basic English phrases, as if they were complex inflected word-forms. When paired with an already existing FST computational morphological model for Plains Cree, these computational tools allow for 1) providing an inflected Plains Cree word-form that roughly matches the meaning and morphosyntactic features of a simple English phrase, and 2) generating a simple English phrase that matches the meaning and morphosyntactic features of an inflected Cree word-form.

For analyzing and generating English phrases, both tools make use the FOMA compiler (Hulden 2009) for Xerox-style finite-state transducer (FST) specifications (Beesley and Karttunen 2003).[1] These FSTs are applied on the English definitions of Plains Cree entries in three bilingual Cree-to-English dictionaries. These dictionaries are 1) *nêhiyawêwin : itwêwina / Cree : Words* (CW: Wolvengrey 2001), 2) *Maskwacîs Dictionary of Cree Words* (MD: Maskwachees Cultural College 2009), and 3) *Alberta Elders' Cree Dictionary* (AECD: LeClaire and Cardinal 2002). For analyzing and generating Plains Cree word-forms, we use the already existing finite-state morphological model (Snoek et al. 2014; Harrigan et al. 2017), which has been compiled with the Helsinki FST compiler (HFST: Lindén et al. 2011).

The first tool noted above provides an alternative to presenting the results of analyzing a morphologically complex Cree word-form by presenting the morphosyntactic analysis tags, whether as is or in relabeled form (into plain English or plain *nêhiyawêwin* labels), as has been the standard previously in intelligent on-line dictionaries (e.g. Arppe et al. 2022: 22-23, 59-61). The second tool provides an alternative to finding an inflected word-form with the appropriate morphosyntactic features by looking up that form in a paradigm table, often quite daunting in their extent (Arppe et al. 2022: 19-24). These tools are already integrated to provide these two functionalities in the on-line Plains Cree – English dictionary, *itwêwina* (`itwewina.altlab.app`) (Arppe et al. 2022; Arppe et al. 2023; see also Appendix E in this paper), built with the open source *morphodict* intelligent dictionary platform, developed in the *21st Century Tools for Indigenous Languages* (21C) project hosted by the Alberta Language Technology Lab (ALTLab).[2]

[1] Source: `https://github.com/giellalt/lang-crk/tree/main/src/fst/transcriptions`

[2] `https://morphodict.readthedocs.io/en/latest/index.html`

Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), pages 48–62
May 4, 2025 ©2025 Association for Computational Linguistics

Nevertheless, one should note that mapping the English features with Cree features, and finding the Cree entry matching the English lexical content requires software modules that are not covered in this paper.[3] As far as we are aware of, *itwêwina* is the only implementation of a combined integration of both simple English phrase analysis and generation in an on-line dictionary, at the same time effectively implementing the only translation system to/from Cree, though in a very restricted form. General machine translation solutions have been developed, of course, for majority and other languages, in particular well-resourced ones. However, the current state-of-the-art approaches that such MT systems rely on require large amounts of parallel corpus data; the closest to this in the North American context that one has come for Indigenous and polysynthetic languages has been for the pairing of Inuktitut and English (Littell et al. 2018; Knowles et al. 2020; Le and Sadat 2020; Microsoft Translator 2021; Caswell 2024).

Originally, the morphosyntactic features that are covered in the above two tools were based on those included in the so-called extended paradigms for nouns and verbs, as specified for the online *itwêwina* dictionary[4]. For nouns, these are based on unpublished complete paradigm layouts provided by Arok Wolvengrey (p.c.); for verbs, these are also largely based on published paradigms provided by Wolvengrey (2011: 393ff, Appendices A and B). For nouns, these extended paradigms include all the possible inflectional features, namely singular, plural, obviative, locative and distributive forms, for the non-possessed word-forms as well as with all possible possessors. For verbs, these extended paradigms include all the possible person and number combinations for subjects, as well as objects, when applicable (only for transitive animate verbs). For all possible subject-object combinations, the most common cases of tense/aspect/mood (expressed by prefixes known as *preverbs*) are included, namely the unmarked case (often referred to as the present tense, usually translated as "s.t. **happens**" or "s/he **does** s.t."), the past (*kî-*, "s/he **did** s.t."), future definite (*ka-*, "s/he **will** do s.t."), future prospective (*wî-*, "s/he **is going to** do s.t."), and the infinitive/irrealis (*ka-* and *ta-*, "**for** s.o. **to** do s.t.") tenses/moods. In addition, all subjunctive, *aka* future conditional forms (translated usually as "**when** s.o. does s.t."), as well as imperative forms in both the immediate and delayed cases are included (translated as "(you) do something **now**" or "(you) do s.t. **later**", respectively).

Later on, the set of morphosyntactic features has been expanded (only for English verb phrase generation) to include, not only those that are relevant for Plains Cree, but also other languages in the Algonquian and Dene language families. This has led to covering negated ("s/he does **not** do s.t.") and progressive forms as well as dual (e.g. "we **both**") and distributed plural (e.g. "**each and every one of us** does s.t.") and indirect object arguments (e.g. "s/he gives s.t. **to us**" or "s/he does s.t. **for us**") for verbs.[5] Nevertheless, one should note that the English phrase types that can be analyzed and generated by the tools presented here are only a small and quite restricted subset of the entire set of possible English constructions, even though the selected subset can be considered as the most common of English simple construction types, with the highest relevance for the most common inflected Cree noun and verb word-forms.

2 Implementation – Analysis

2.1 Analysis of English verb phrases

The analysis (and generation) of English verb phrases, relies on two factors. Firstly, English personal pronouns indicating subjects, objects, and possessors are mostly distinct from each other, with the exception of the second person *you* and third person neuter *it*. This allows for the identification of these arguments and their relevant syntactic roles in simple phrases with a single predicate; when one or more of such verbal arguments can be identified, that is interpreted to indicate a verbal phrase to be matched with a Plains Cree verb form.

[3] https://github.com/UAlbertaALTLab/morphodict/tree/main/src/morphodict/phrase_translate

[4] https://github.com/UAlbertaALTLab/morphodict/tree/main/src/morphodict/paradigm/layouts

[5] The English verb phrase generator can be extended to other languages, provided that their morphosyntactic features can be mapped to available English auxiliary constructions, and the English dictionary definitions follow a templatic structure similar to the three Cree lexical resources referred to in this paper. E.g. for Tsuut'ina (ISO: srs; Dene language family), its Imperfective, Perfective, and Progressive verbal aspects can be mapped to the English Future Definite, Past, and Present Progressive tenses, respectively, and the Repetitive subaspect to the English Repetitive adverbial construction (with "again and again"). We have initially explored this with encouraging results.

Tense/aspect/modality feature(s)	Initial zone	Subject/ Existential	Predicate zone	Object/ Reflexive	Final zone
Present		**we**	help>0	**you**	
Past		**we**	help>ed	**you all**	
Future+Definite		**we**	will help	**you**	
Future+Prospective		**we**	are going to help	**you all**	
Imperative+Immediate / Delayed	let	**you and us**	help	**him**	now/later
Imperative+Immediate+Negation	do not let	**you and us**	help	**her**	now
Infinitive	for	**us**	to help	**you**	
(Future) Conditional	when	**we**	help	**you all**	
Present (Existential)		**there**	is	**light**	
Present+Negation (Existential)		**there**	is not	**light**	
Present (Copula)		**it**	is red		
Present+Negation		**we**	do not help	**you**	
Present (Copula)		**we**	are ready		
Present+Negation (Copula)		**we**	are not ready		

Table 1. Examples of various English verb constructions, split into the various templatic zones with the help of subject/existential and object (reflexive) markers (in blue). The parts of the Predicate zone indicating Tense/Mood are marked in red. Analyses of the above verb phrases are shown in Appendix A.

Secondly, combined with the relatively strict word order of English, these subject and object personal pronouns allow for the partitioning of English phrases into a *templatic structure* with an initial, predicate, and final zone.

In the *predicate* zone, immediately after the subject pronoun (or existential marker) and before the object pronoun, when available, usually the first word is either 1) an auxiliary verb (e.g. *will* or *do*, *does*, or *did,* or the copula *am/is/are/was/were*, by itself or as part of an auxiliary phrase, e.g. *am/are/is/was/were going to*), or 2) a finite verb form (e.g. *help*, *helps*, or *helped*). This enables the determination of the tense, aspect, and modality of a phrase. The *initial* zone, preceding the first appearance of a personal pronoun indicating the subject, enables the identification of imperative/permissive constructions (and their negation), future conditional, and infinitival constructions, indicated by the initial elements *let*, *when*, and *for*, respectively. Generally, the initial personal pronoun would be in subject form, e.g. "**I** do s.t." or "when **we** do s.t.", but in the case of an initial zone preposition *for* or the auxiliary verb *let*, the initial personal pronoun will take the object form, in e.g. "**for me** to do s.t." or "**let us** do s.t.". The final zone is mainly used for the identification of the immediate or delayed subtypes of imperative constructions introduced above, as well as repetitive forms, indicated by the adverb constructions *again and again* or *repeatedly*. Examples of elements in these zones for the verb phrase template, with their linguistic analyses for tense/aspect/modality, are provided in Table 1 above.

In this current implementation, personal pronouns and certain other nominal expressions (i.e. 'someone' and 'people' for unspecified subjects in Plains Cree) which are identified as arguments are converted into flag diacritics, of the P-flag type that only sets the value of a flag diacritic variable, without checking for any constraints. These flags will then each represent a subject, direct object, indirect object, or reflexive feature. An example of this for the regular subject pronouns is given below in (1).

One may firstly notice that for certain multiword subject expressions, such as ones corresponding to the Plains Cree *further obviative* feature (sometimes referred to as the 5th person), alternative versions are recognized. Secondly, longer subject constructions are attempted to be matched before shorter ones in separate regular expressions that are then composed together, to ensure only a single maximal match. Thirdly, the targeted subject constructions are expected to be demarcated with boundary characters defined as the regular expression `Bx`, consisting of phrasal punctuation characters, the space character, and the input boundary (both initial and final). Finally, we apply a convention where the second person pronoun "you" is interpreted as singular (`+2Sg`) when occurring by itself, and, when occurring with "all" as "you all", as the (exclusive) plural (`+2Pl`).

(1)
```
regex [ {yet another} | {yet others} |
{he/she/they over there} | {he or she or
they over there} | {she or he or they over
there} | {he, she, or they over there} |
{she, he, or they over there} ->
"@P.subject.5Sg/Pl@" || Bx _ Bx ]
```

```
.o. [ {another} | {others} | {he over
there} | {she over there} | {they over
there} | {he/she/they} | {he or she or
they} | {she or he or they} | {he, she, or
they} -> "@P.subject.4Sg/Pl@" || Bx _ Bx
]
.o. [ {it} -> "@P.object.0Sg@" || Bx _ "
" .#. ]
{you all} -> "@P.subject.2Pl@" || Bx _ Bx
,,
{you} -> "@P.subject.2Sg@" || Bx _ Bx ,,
...
{it} -> "@P.subject.0Sg@" || Bx _ Bx ,,
... ];
define Subject
```

Furthermore, matched elements in the initial zone are converted into P-type flag diacritics representing the associated modality features. In the predicate zone, matched auxiliary verbs and auxiliary phrases as well as the negative adverb *not* are similarly converted into P-type flag diacritics. Sometimes, the interpretation of the subject pronouns is combined with a scrutiny of relevant elements in the initial zone, a fragment of which is exemplified below in (2) for the infinitival phrases, initiated with "for …".

(2)
```
regex [ ...
[ {for you all to} -> "@P.subject.2Pl@"
"@P.tense.Inf@" || .#. _ " " ,,
{for you and us to} | {for you and we to}
| {for you and me to} | {for us and you
to} | {for we and you to} | {for me and
you to} -> "@P.subject.21Pl@"
"@P.tense.Inf@" ,
... ] .o.
[ {for me to} -> "@P.subject.1Sg@"
"@P.tense.Inf@" || .#. _ " " ,,
{for you to} -> "@P.subject.2Sg@"
"@P.tense.Inf@" || .#. _ " " ,,
... ];
define ReverseSubject
```

All these conversions result in removing the function words from the output, and leaving only the lexical content words that will be used in the subsequent English-to-Cree search. In conjunction with this, finite non-auxiliary verb forms that occur in the very beginning of the predicate zone are analyzed as to their tense (present, past, or infinitive) and an appropriate P-type flag diacritic is added, with the verb standardized into the bare infinitive form. Here, for irregular verbs a set of 155 tuplets of past tense and the corresponding infinitive forms are used (specified as `input:output` pairs in a LEXC file); all other, regular verb forms are analyzed using regular expressions recognizing the regular suffixes *-(e)s*, *-(e)d*, and *-ing*, and removing such suffixes to provide the bare infinitive/stem form.

Overall, the English verb phrase analyzer consists of a number of constituent regular expressions such as the ones for `Subject` and `ReverseSubject` presented above, as well as regular expressions for standardizing the input. These are then composed together in a specific order, which is intended to result in appropriate matches. For verb phrases, the ordering is presented below (3). For the most part, this is self-explanatory, in particular for identifying the various argument elements and analyzing the predicate zone. Nevertheless, one first specifies 1) the alphabet on which the English phrase analyzer operates on (which applies for both verbal and nominal phrase types), followed by 2) the insertion of an additional final space to satisfy the boundary requirements for many of the argument-marking component regular expressions, and then 3) the standardization of certain forms as well as masking apostrophes in possessed forms, before 4) undoing contractions such as "don't" as "do not", followed by 4) the resurrection of the apostrophe for possessed word-forms.

(3) Letters+ → InsertFinalSpace → Standardization → Contractions → ResurrectApostrophe → Negation → Past → ReverseSubject → Reciprocal → Subject → Reflexive → Object → 1DuO → 2SgO → 2DuO → 2PlO → 3SgO → 3DuO → 4SgPlO → 5SgPlO → 0SgO → XO → XPlO → Tense → Imperative → Progressive → Repetitive → VerbClass → VerbAnalysis

The next constituent regular expressions from `Negation` through `Repetitive` should be self-explanatory, except for the rule sequence `1DuO → 2SgO → 2DuO → 2PlO → 3SgO → 3DuO → 4SgPlO → 5SgPlO → 0SgO → XO → XPlO`, which concerns the reinterpretation of certain ambiguous argument types as an object, if the argument in question is already preceded in the phrase by a subject argument. Furthermore, the next-to-last regular expression uses the occurrence or absence of identified subject and object arguments and their semantic types to determine the Plains Cree verbal part-of-speech corresponding to the English phrase, based on transitivity of the predicate and animacy of the arguments (as an II, AI, TI, or TA type of verb). This allows for matching the correct type of Cree verb entry for the English phrase.

Finally, all the aforementioned P-type flag diacritics determine the generation of corresponding analysis tags. In fact, all the theoretically possible analysis combinations are generated, but the correct analysis is filtered by matching R-type (for *Require*) flags; the resultant tags are output after the standardized lexical content words using the final regular expression `VerbAnalysis`. A simplified example of the structure of the input and output and the intervening intermediate form with the P-type flags is shown in (4a). After that, the English analysis output is converted into input matching the tags used by the Cree generator FST, resulting in an inflected Cree word-form approximating the original English phrase (4b).

```
(4a) I am going to see you all
→ @P.subject.1Sg@ @P.tense.Fut@ see
@P.object.2Pl@
→ see +V+TA+Fut+1Sg+2PLO
```

```
(4b) → PV/wî+wâpamêw+V+TA+Ind+1Sg+2PLO
→ kiwî-wâpamitinâwâw
```

One should note that mapping the English tags into Cree ones and finding the Cree entry matching the English lexical content makes use of a *morphodict* software module and specifications that are not covered here.[6]

2.2 Analysis of English noun phrases

Similar to verbs, the analysis (and generation) of English noun phrases relies on a relatively fixed word order for the nominal head and locative prepositions, and otherwise on certain identifiable modifiers. The maximally complex noun phrase that we try to analyze is limited by what can be mapped to an inflected Plains Cree nominal word-form, examples of which are presented in Table 2.

The noun which is the head of the noun phrase is expected to be found at its end, identifiable by an immediately following punctuation character understood as separating multiple noun phrases, or a string final boundary; A Plains Cree noun can express either the singular or plural number or obviation, but not both. For the purposes of English noun phrase analysis, plural number can be expressed either morphologically with the suffix –*(e)s* on the final word interpreted as a noun, or with several attributes ("many", "few", "several", or "couple of") occurring as the first element of the noun phrase, while initial articles or indefinite pronouns ("a", "an", or "one") are used to denote a singular number. Obviation can be expressed by adding the phrase "over there" after the final noun, which will override any preceding expression of number. In addition, a noun phrase can be initiated with an optional locative preposition, either "in" or "on", or the optional *distributive* locative preposition "among", both of which in Plains Cree are again mutually exclusive with singular or plural number or obviation. In addition, one can indicate an optional possessor with English possessive pronouns, or certain nouns ("someone's" or "people's") standing in for the unspecified possessor. Furthermore, one can optionally signal a diminutive form by using any of the modifying attributes "little", "lesser", "smaller", or "younger", though currently this is not used to create corresponding Cree noun forms.

Locat-ive	**Possess-or**	**Num-ber**	**Dimin-utive**	**Noun head**	**Obv-iation**
		one		**bear**	
		many		**bear>s**	
	another			**bear**	over there
	my	one		**book**	
	my		little	**book**	
in		a		**book**	
among				**bear>s**	
in	my	many		**book>s**	
	my other		little	**tree**	over there

Table 2. A sample of English noun phrases corresponding to Plains Cree noun forms, and their templatic structure. Analyses are shown in Appendix B.

Similar to verbs, the English noun analyzer consists of a number of constituent regular expressions that are applied in a sequence shown below (5). The first regular expression `NounPl2Sg` analyzes the final word of the phrase expecting that to be a noun, and if this appears to be a plural form, either having either of the regular *-s* or *-es* suffixes or being one of the enumerated 79 pairings of irregular plural forms with their singular counterparts, a P-type flag indicating plural number is affixed and the plural form replaced by the corresponding singular form. This is followed by two regular expressions `VerbPhrase` and `NounPhrase`, used to disallow a noun phrase analysis if any of the personal pronouns representing subjects or objects are present. After this, a succession of regular expressions converts various function words into corresponding flags indicating number, obviation,

[6] For the tag mappings between Cree and English, see: https://github.com/UAlbertaALTLab/morphodict/tree/main/src/crkeng/resources/phrase_translate

location, diminutivization, and possession, where their actual order in the noun phrase has little role.

As one might well write a phrase combining a locative preposition and any of the markers of number or obviation, as these are converted into P-type flag-diacritics, only the value of the last flag will remain in effect. An exception is implemented with the regular expression `DistrVsPl`, with plural flags deleted after an initial distributive preposition, since that construction is usually translated into English with a plural noun form, e.g. "among the Americans" for *kihci-môhkomâninâhk*, even though the corresponding Cree noun form is underspecified in terms of its number.

The next-to-final regular expression `NounClass` outputs a tag (`+N`) indicating a noun analysis, whereas the final regular expression `NounAnalysis` outputs analysis tags corresponding to the P-type flag-diacritics generated earlier. Again, all the theoretically possible combinations of noun analysis tags are output, with R-type flags filtering the appropriate correct analysis. One should note that one cannot currently map an English noun phrase to the two Plains Cree animacy types, as English has no such morphosyntactically expressed distinction.

(5) NounPl2Sg → VerbPhrase → NounPhrase
→ NumberObvLocDistr → Diminutive → Possession
→ DistrVsPl → NounClass → NounAnalysis

Importantly, the noun analyzer described here is treated as disjunct to the verb analyzer described earlier, resulting in a single finite-state analyzer for English phrases. The (uncompressed) size of the resultant compiled FOMA binary file is 24.4 MB.

3 Implementation – Generation

3.1 English verb phrase generation

The easy identifiability of the subject and object personal pronouns and the associated zones in a simple English verb phrase, as described above in the analysis of such phrases to present a general *templatic* structure, also enables the manipulation of the subject and object pronouns as well as the tense, aspect, and modality of the predicate in such a phrase, allowing for the generation of new phrases matching the morphosyntactic features expressed by an inflected Cree word-form. Indeed, an examination of English definitions in the three bilingual Plains Cree-to-English dictionaries of which we have access to their content in electronic form shows that these all follow particular structures and expressions that support such manipulation. Examples from CW, MD, and AECD for the three Cree entries, *wîcihêw, nisitohtam*, and *mispon*, are given below in (6-8).

(6) *wîcihêw* (to help s.o.)
- **s/he** helps **s.o.**, **s/he** assists **s.o.** (CW)
- **s/he** provides welfare to **s.o.** (CW)
- **s/he** assists **s.o.** in childbirth, **s/he** serves as a midwife to **s.o.** (CW)
- **He** helps **him**. (MD)
- **He** aids **him**. (MD)
- **They** help **him**. (MD)
- **s/he** assists **her/him** or **them** (AECD)
- **s/he** participates (AECD)

(7) *nisitohtam* (to understand [s.t.])
- **s/he** understands (CW)
- **He** understands (MD)
- **s/he** understands **s.t.** (CW)
- **s/he** understands **it** (AECD)

(8) *mispon* (to snow)
- **it** snows, **it** is snowing, **there** is falling snow, **there** is a snowfall (CW)
- **It** is snowing. (MD)
- **it** snows, or **it** is snowing (AECD)

While the indication of subjects and objects above, in red and blue, respectively, varies between the three dictionary sources, with "s/he" in CW and AECD vs. "He" in MD for animate subjects, and "it" or "there" throughout for the inanimate subjects/existential marker, and "s.o." in CW vs. "her/him" or "them" in AECD vs. "him" in MD for animate objects (whether for direct or prepositional) and "s.t." in CW vs. "it" in AECD and MD for inanimate objects, the use of specific conventions, e.g. "s/he" or "him/her", in fact makes their identification easier than for the English phrase analysis. Also, the main finite verbs in each of the definitions, underlined above, practically always immediately follow the subject marker, which substantially facilitates their manipulation and/or the insertion of various auxiliary verb constructions immediately after the subject marker, sometimes in conjunction with an additional element appended in the initial zone preceding the subject, to convey a broad range of tense, aspect, and modality features. Finally, in all three sources both commas and semi-colons are almost invariably used only to delineate different senses, further clarifying where the initial zone/subject is to be found in any of the definitions above.

(9) Input → itNonSubjectMarking → Standardization
→ VerbInflection2 → ModalAuxiliary

→ Tag2FlagConversion → VerbInflection → IrregularVerbs → AuxiliaryPhrase → BeAuxiliaryPhrase → PhrasePrefix → PhraseSuffix → Relativization → BeInflection → IObjectConversion → SubjectConversion → ObjectConversion → ReflexiveConversion → PossessiveConversion → itNonSubjectUnmarking → EngInf2Inf → EngPrt2Prt → EngPrf2Prf → EngProg2Prog → EngPhon → RemoveSuffixBoundary

The actual implementation of the manipulation of the English definitions again consists of a sequence of often quite complex regular expressions, enumerated above in (9). The application of these regular expressions depends on the assumed templatic structure of the base definition phrase, as exemplified earlier above in (6). In general, at various positions in the template the regular expressions either 1) identify elements as particular arguments or modifiers with place-holder markers which are eventually converted into the desired target form, or 2) insert elements to convey the desired tense, modality, and polarity.

As the very first thing, various tags indicating the desired manipulation, in terms of the subject and direct object as well as tense, aspect, and other modalities including negation, are expected to precede the English definition that is to be manipulated. These tags are converted into P-style flag-diacritics that can be used to select and filter the desired manipulation. The order and optionality of the full set of possible tags is enumerated in the first regular expression, named `Input`, shown here in (10) below. A detailed overview of the possible tags available for English verb phrase generation, with their combinatorics and exponents, is presented in Appendix C.

(10)
```
regex [ (Clause | Modality) [ TenseAspect | Auxiliary ] Subject (DirectObject) (IndirectObject) (Progressive) (Repetitive) (Negation) Letters+ ] ;
define Input
```

After this, the regular expression named `Standardization` is used to identify and replace with unambiguous multicharacter symbols the various argument markers for the subject or existential "there", the possessive and reflexive markers that should correspond with the subject, the possible direct object and indirect object markers, and the copula verb "is", so that they cannot be confused with the subsequent manipulations. This is followed by a sequence of regular expressions from `VerbInflection2` through `BeInflection`, which either 1) insert various auxiliary verbs or longer auxiliary constructions, sometimes in co-ordination with an prepended element in the initial zone before the subject/existential marker or at the end of the phrase (before the sense-demarcating punctuation), or 2) convert the finite copula verb or finite regular verb, occurring immediately after the subject marker, into the appropriate forms which are needed for agreement with the auxiliary verb constructions, i.e. present or past tense, present (progressive) or past participle, or the bare infinitive form. Importantly, the entire set of these alternative constructions are always generated, but they are marked with sets of R-type flag diacritics that indicate which initial tag-based P-type flag diacritics they are allowed/required to co-occur with, resulting in (ideally) only one of the generated manipulations getting filtered for final output. To simplify encoding, many P-type and R-type flags, as well as D-type flags (for disallowing contexts) are grouped into named sets that explicitly enumerate the conditions where the particular variant English verb-form is the appropriate one.

Auxiliary Construction	**Flags**
wants to	@R.tense.Int@ @D.neg@ @D.prog@ RsubjectPrs3Sg
want to	@R.tense.Int@ @D.neg@ @D.prog@ DsubjectPrs3Sg
does not want to	@R.tense.Int@ @R.neg.Neg@ @D.prog@ RsubjectPrs3Sg
do not want to	@R.tense.Int@ @R.neg.Neg@ @D.prog@ DsubjectPrs3Sg
wants to be	@R.tense.Int@ @D.neg@ @R.prog.Prog@ RsubjectPrs3Sg
want to be	@R.tense.Int@ @D.neg@ @R.prog.Prog@ DsubjectPrs3Sg
does not want to be	@R.tense.Int@ @R.neg.Neg@ @R.prog.Prog@ RsubjectPrs3Sg
do not want to be	@R.tense.Int@ @R.neg.Neg@ @R.prog.Prog@ DsubjectPrs3Sg

Table 3. The 8 possible realizations of the Future Intentional auxiliary phrase, with associated flags.

Furthermore, in a step beyond the verb phrase analyzer, the various auxiliary verbs and constructions are organized as completely written out sets of constructions following 4-8 patterns, depending on the possible variation arising for each tense/aspect/modality feature and negation; these constructions are then inserted as single "pre-fabricated" chunks between the first subject-marking pronoun and the immediately following predicate zone. Due to diverging patterns between copular and regular finite verbs, two sets of constructions are specified. Crucially, each construction determines what form of the finite verb (or copula) is required (to follow the auxiliary construction). This templating approach turned out absolutely necessary in order to allow for the efficient maintenance of the verb phrase generator, and it has also proven to enable the easy addition of new modalities, in comparison to the jungle of flag-diacritics that arose from initially trying to insert the exponents of various features individually one by one. An example of the set of constructions for various variants of the *future intentional* modality is shown in Table 3 above.

After the generation and insertion of all the possible variants, the various subject, object, and other multicharacter markers are converted with the regular expressions from `IObjectConversion` through `itNonSubjectUnmarking` into the pronouns or nouns specified in the original tags. Then, the regular expressions from `EngInf2Inf` through `EngProg2Prog` provide the correct forms for irregular or regular English verbs occurring in the CW dictionary, while the subsequent regular expression `EngPhon` deals with the orthophonemic variation in the case of all the remaining regular verbs, and the final regular expression `RemoveSuffixBoundary` cleans up the results. An example of the results of English phrase generation, with a first person singular subject and a third person plural object (when applicable), in the future prospective tense, is shown below in (11-13) for the English definitions provided above (in 6-8), with subject/object/existential markers in blue and the inserted auxiliary construction in red.

The FOMA-compiled English verb phrase generator reaches 30.6 MB in (uncompressed) size. One should again note that the full implementation involves a *morphodict* code module, with mappings between Plains Cree word-form analysis tags to English phrase generation tags, which is not discussed here (but see Footnote 7 above).

(11) *niwî-wîcihâwak* (to help s.o.)

- **I** am going to help **them**, **I** am going to assist **them** (CW)
- **I** am going to provide welfare to **them** (CW)
- **I** am going to assist **them** in childbirth, **I** am going to serve as a midwife to **them** (CW)
- **I** am going to help **them**. (MD)
- **I** am going to aid **them**. (MD)
- **I** am going to help **them**. (MD)
- **I** am going to assist **them** (AECD)
- **I** am going to participate (AECD)

(12) *niwî-nisitohtên (to understand [s.t.])*

- **I** am going to understand (CW) (MD)
- **I** am going to understand **something** (CW) (AECD)

(13) *wî-mispon* (to snow)

- **it** is going to snow, **it** is going to be snowing, **there** is going to be falling snow, **there** is going to be a snowfall (CW)
- **It** is going to be snowing. (MD)
- **it** is going to snow, or **it** is going to be snowing (AECD)

3.2 English noun phrase generation

Similar to the verbs, the structure of the English definitions in the three Plains Cree-to-English dictionaries is convergent, with commas/semicolons used to distinguish different senses. But in contrast to verbs, our task is both simpler and more difficult, in that we only need to identify a single anchor word in each noun phrase/sense, namely the noun head, but there is no unambiguous marker word we can rely on throughout; while empirical investigation of the definitions indicate this to be mostly either the word immediately preceding a postmodifying prepositional or relative phrase, or otherwise the final word preceding the sense-demarcating punctuation character or the end of the definition, not all prepositions initiate a postmodifying phrase. An example of English definitions for *okimâw* "chief" is provided below in (14), with the head noun marked underlined in **bold-face**.

(14) *okimâw* "chief":

- **chief**, **leader**, head **person**, **man** of high position (CW)
- **king** (CW)
- **boss** (CW)
- one's **superior** (CW)
- **manager** (CW)
- A **chief**. (MD)
- A **man** in high position. (MD)
- a **leader** on a job site, i.e.: a boss (AECD)
- government **leader**, **manager** (AECD)

Anyhow, the generation of English noun phrases that can be matched with the inflectional features available for Plains Cree noun forms is organized similar to the verb phrase generation described above, but it is overall substantially simpler, and currently has not been partitioned into named constituent regular expressions. One needs only specify one among the mutually exclusive features for either singular or plural number, obviation, locative or distributive form, followed by an optional possessor and optional diminutivization. A detailed overview of the possible tags available for English noun phrase generation with their exponents is presented in Appendix D.

As with the verbs, the tags specifying these features precede the definition to be manipulated, and are first converted into P-type tags. Then, any initial articles or possessive pronouns in the original English definition are removed. After this, an optional modifier "little" indicating a diminutive form can be prepended to the remaining noun phrase, which in turn can be prepended with an optional possessive pronoun, followed by the pronoun "other" in the case of an obviative form. Finally, the possible alternatives among the number/obviation/location complex are added: either the locative or distributive preposition, "in" or "among", is prepended at the very beginning, or an indefinite article for singular number (which combines into "another" in the case of obviation), or the plural suffix *–(e)s* is added to the word at the end of the noun phrase marked by a final string boundary or punctuation separating senses, or if the noun phrase ends with a postmodifying phrase starting with "of", "for", "with", "among", or "who", then to the word immediately preceding these prepositions or relative pronoun. For some 88 irregular English plural nouns the appropriate forms are enumerated, while for the remaining pluralized nouns regular orthophonemic rules are applied. As with the verbs, the complete set of all possible outcomes are generated with associated R-type flags, which allow for the selection of the one desired noun phrase which matches the P-type flags specified by the initial tags. An example of creating plural first-person-singular possessed forms of the definitions above is shown here in (15) (with the incorrect generations ~~struck-through~~). The FOMA-compiled English noun phrase generator reaches 56.9 MB in (uncompressed) size.

(15) *nitokimâmak* "chief":

- my **chiefs**, my **leaders**, my head **persons**, my **men** of high position (CW)
- my **kings** (CW)
- my **bosses** (CW)
- ~~my one's **superiors**~~ (CW)
- my **managers** (CW)
- my **chiefs**. (MD)
- ~~my **man** in high positions~~. (MD)
- ~~my **leader** on a job sites~~, i.e.: my bosses (AECD)
- my government **leaders**, my **managers** (AECD)

4 Evaluation

To evaluate the English phrase generation, focusing on verbs, with greater complexity, a combination of the extended Plains Cree layouts, a Plains Cree corpus, and the English definitions in the three dictionaries was used to attempt to generate 2253 English phrases corresponding to the Cree morpho-syntactic features for each word-form cell in all these verb paradigms. A quantitative scrutiny showed the English phrase generation was able to generate a manipulated phrase for 2151 (95.5%) of the verb cells, taking 0.457s on a MacBook Pro with an Apple M4 Max CPU. A comprehensive manual evaluation of the generated phrases is under way, but preliminary results for the first 1500 phrases indicate that 1252 (83.5%) are fully well-formed. For evaluating English phrase analysis, we used the aforementioned generations to create 5430 simple English verb phrases, which were then run through the analyzing transducer. For 4268 (78.6%) of these phrases, the transducer provided exactly the same set of morphosyntactic features as were used to generate the phrase.

A qualitative evaluation of the remaining unsatisfactory behavior suggests that this is due to co-ordinated predicate constructions ("Xs and/or Ys") not yet covered by the verb phrase generator, structural ambiguity of some high-frequency verbs (e.g. "lie/lied" vs. "lie/lay"), and missing some orthographical variants of the subject and object markers in the English definitions in the three dictionaries. Another point of improvement concerns the identification and removal of some forms of parenthetical content in the English definitions, usually marked with bracketing, which can confuse the phrase transductors, if missed.

5 Conclusion

This paper presented rule-based finite-state transducers for analyzing and generating simple English phrases, matching the most common inflected Plains Cree verb and noun word-forms. These transducers have been incorporated in the on-line intelligent Plains Cree-English dictionary, *itwêwina*, using the *morphodict* platform (see Appendix E for example screenshots). Combined with a matching Plains Cree morphological transducer and mappings between the English and Plains Cree features, this in effect results in restricted machine translation between the two languages. The morphosyntactic features covered by these English phrase transducers have been extended beyond Cree to cover ones apparent in other Algonquian and Dene languages. While not explored in detail in this paper, this can in principle enable the implementation of the same functionalities for other similar languages.

Limitations

The English phrases that the tools presented in this paper can analyze and generate are restricted to constructions roughly matching the inflected Cree word-forms contained in so-called extended paradigms. One must keep in mind that these recognized constructions are only a very small subset of all possible constructions in English; while the most common Cree word-forms are covered, many rarer inflected word-forms are not. Perhaps more importantly, the generated English phrases should be considered only as rough approximate translations of the corresponding Cree word-forms, and should **not** be used as a replacement for consulting fluent speakers for translations.

The compiled data structures resulting from the finite-state approach employed in this paper, when combining both the analysis/generation of English phrase structure and of certain English word-forms within these constructions, can end up being prohibitively large (in hundreds of megabytes or even bigger), which will be challenging to integrate within other applications, and may be slow to look up. Features that are necessary, such as keeping track of already seen elements (memory) and the precise identification of word and phrase boundaries, are cumbersome to implement in finite-state specifications, requiring flag-diacritics that are notoriously difficult to parse and debug. Based on some preliminary trials, procedural coding approaches (e.g. with Python) would appear to provide an alternative that can implement all the desired features incorporated in the finite-scale models discussed in this paper, and beyond, while at the same time being sufficiently fast and not exploding required memory.

Ethics Statement

The on-line Plains Cree-English dictionary described in this paper has been developed in order to support the explicit objectives of Cree language communities to support their language instruction, maintenance, and revitalization activities. The functionality presented in this paper has been fine-tuned based on feedback from various individuals in Cree-speaking communities.

Acknowledgments

The English phrase analysis and generation tools presented in this paper are greatly facilitated by the systematic structure of the English definitions in the three Plains Cree-English dictionaries that we have received access to, courtesy of Dr. Earle Waugh (AECD), the Maskwacîs Education Schools Commission (MD), and Dr. Arok Wolvengrey (CW). I am grateful to all their authors and editors for their consistency, and in particular Dr. Wolvengrey, who has also proposed the standard English translation templates for the most common inflected Cree word-forms, forming the basis for the generation of the corresponding English phrases. The initial idea for connecting inflecting Cree word-forms directly with matching English phrases was suggested to me by Dr. Jordan Lachler, while the flag-diacritic approach which allows for keeping the size of the compiled transducers reasonable was suggested and explained to me by Dr. Miikka Silfverberg. The practical integration of the English phrase analyzers as a functionality within *itwêwina* and *morphodict* was implemented and improved in various stages by Andrew Neitsch, Jolene Poulin, and Dr. Felipe Bañados Schwerter. The details of the implementation have evolved based on feedback from Cree and other learners and instructors in the communities and the academia. The current though yet on-going quantitative evaluation of the transducers is based on manual scrutiny by Alexis Ibarra. Finally, this work has been supported by the SSHRC Partnership Grant *21st Century Tools for Indigenous Languages* (895-2019-1012).

References

Antti Arppe, Andrew Neitsch, Daniel Dacanay, Jolene Poulin, Daniel Hieber, and Atticus Harrigan. 2023. Finding words that aren't there: Using word embeddings to improve dictionary search for low-resource languages. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 144–155, Toronto, Canada. Association for Computational Linguistics.

Antti Arppe, Jolene Poulin, Eddie Antonio Santos, Andrew Neitsch, Atticus Harrigan, Katherine Schmirler, and Arok Wolvengrey. 2022. Towards a morphologically intelligent and user-friendly on-line dictionary of Plains Cree – next next round. Presentation at the *54th Algonquian Conference*, Boulder, Colorado, October 20-23, 2022. https://altlab.ualberta.ca/wp-content/uploads/2023/03/itwewinaAC54Oct2022.pptx.pdf

Isaac Caswell. 2024. Google Translate learns Inuktut. https://blog.google/intl/en-ca/company-news/technology/google-translate-learns-inuktut/. Accessed: 2024-11-05.

Atticus Harrigan., Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud, and Arok Wolvengrey. 2017. Learning from the computational modelling of Plains Cree verbs. *Morphology*, 27, pages 565–598.

Mans Hulden. 2009. Foma: a finite-state compiler and library. *Proceedings of the Demonstrations Session at EACL 2009*, pages 29–32.

Rebecca Knowles, Darlene Stewart, Samuel Larkin, and Patrick Littell. 2020. NRC Systems for the 2020 Inuktitut-English News Translation Task. In *Proceedings of the Fifth Conference on Machine Translation*, pages 156–170. Association for Computational Linguistics.

Tan Ngoc Le and Fatiha Sadat. 2020. Low-Resource NMT: an Empirical Study on the Effect of Rich Morphological Word Segmentation on Inuktitut. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 165–172. Association for Machine Translation in the Americas.

Nancy LeClaire and George Cardinal, G (compilers); Earle H. Waugh (editor). 2002. *Alberta Elders' Cree Dictionary / alperta ohci kehtehayak nehiyaw otwestamâkewasinahikan*. University of Alberta Press, Edmonton, Alberta.

Krister Lindén, Erik Axelson, Sam Hardwick, Tommi A. Pirinen, and Miikka Silfverberg. 2011. HFST—framework for compiling and applying morphologies. *Systems and frameworks for computational morphology: Second International Workshop, SFCM 2011*, Zürich, Switzerland, August 26, 2011. Proceedings 2, pages 67–85.

Patrick Littell, Anna Kazantseva, Roland Kuhn, Aidan Pine, Antti Arppe, Christopher Cox, and Marie-Odile Junker. 2018. Indigenous language technologies in Canada: Assessment, challenges, and successes. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2620–2632, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Maskwachees Cultural College. 2009. *Maskwacîs Dictionary of Cree Words / Nêhiyaw Pîkiskwêwinisa*. Maskwacîs, Alberta.

Microsoft Translator. 2021. Inuktitut is now available in Microsoft Translator! *Microsoft Translator Blog.* https://www.microsoft.com/enus/translator/blog/2021/01/27/inuktitutis-now-available-in-microsofttranslator/. Accessed: 31-03-2025.

Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2014. Modeling the noun morphology of Plains Cree. *Proceedings of the 2014 workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL)*, Baltimore, Maryland, pages 34–42.

Arok Wolvengrey (editor) 2001. *nêhiyawêwin: itwêwina / Cree: Words*. Canadian Plains Research Center, Regina, Saskatchewan.

Arok Wolvengrey. 2001. *Semantic and pragmatic functions in Plains Cree syntax*. LOT dissertation series 268, LOT, Utrecht, the Netherlands.

Appendix A. Example analyses of English verb phrases in Table 1.

Initial zone + Subject/Existential + Predicate zone + Object/Reflexive + Final zone	**English analysis: Lexical**	**English analysis: Morphosyntax**	**Cree FST generation tags**	**Cree verb form**
we help>0 **you**	help	+V+TA+1Pl+2SgO	wîcihêw+V+TA+Ind+1Pl+2SgO	kiwîcihitinân
we help>ed **you all**	help	+V+TA+Prt+1Pl+2PlO	PV/kî+wîcihêw+V+TA+Ind+1Pl+2PlO	kikî-wîcihitinân
we will help **you**	help	+V+TA+Def+1Pl+2SgO	PV/ka+wîcihêw+V+TA+Ind+1Pl+2SgO	kika-wîcihitinân
we are going to help **you all**	help	+V+TA+Fut+1Pl+2PlO	PV/wî+wîcihêw+V+TA+Ind+1Pl+2PlO	kiwî-wîcihitinân
let **you and us** help **him** later	help	+V+TA+Del+21Pl+3SgO	wîcihêw+V+TA+Imp+Del+12Pl+3SgO	wîcihâhkahk
do not let **you and us** help **her** now	help	+V+TA+Imm+21Pl+3SgO+Neg	namôya+Ipc wîcihêw+V+TA+Imp+Imm+12Pl+3SgO	namôya wîcihâtân
for **us** to help **you**	help	+V+TA+Inf+1Pl+2SgO	PV/ka+wîcihêw+V+TA+Cnj+1Pl+2SgO	ka-wîcihitâhk
when **we** help **you all**	help	+V+TA+Cond+1Pl+2PlO	wîcihêw+V+TA+Fut+Cond+1Pl+2PlO	wîcihitâhki
there is **light**	is light	+V+II+0Sg	kîsikâw+V+II+Ind+3Sg	kîsikâw
there is not **light**	is light	+V+II+0Sg+Neg	namôya+Ipc kîsikâw+V+II+Ind+3Sg	namôya kîsikâw
it is red	is red	+V+II+0Sg	mihkwâw+V+II+Ind+3Sg	mihkwâw
we do not help **you**	help	+V+TA+1Pl+2SgO+Neg	namôya+Ipc wîcihêw+V+TA+Ind+1Pl+2SgO	namôya kiwîcihitinân
we are ready	is ready	+V+AI+1Pl	kwêyâtisiw+V+AI+Ind+1Pl	nikwêyâtisinân
we are not ready	is ready	+V+AI+1Pl+Neg	namôya+Ipc kwêyâtisiw+V+AI+Ind+1Pl	namôya nikwêyâtisinân

Appendix B. Example analyses of English noun phrases in Table 2.

Number / Locative + Possessor + Diminutive + Noun head + Obviation	**English analysis: Lexical**	**English analysis: Morpho-Syntactic**	**Cree FST generation tags**	**Cree noun form**
one **bear**	bear	+N+Sg	maskwa+N+A+Sg	maskwa
many **bear>s**	bear	+N+Pl	maskwa+N+A+Pl	maskwak
bear over there	bear	+N+Obv	maskwa+N+A+Obv	maskwa
my one **book**	book	+N+Px1Sg+Sg	masinahikan+N+I+Px1Sg+Sg	nimasinahikan
my little **book**	book	+Dim+Px1Sg+Pl	masinahikan+N+I+Der/Dim+N+I+Px1Sg+Pl	nimasinahikanisa
in a **book**	book	+N+Loc	masinahikan+N+I+Loc	masinahikanihk
among **bear>s**	bear	+N+Distr	maskwa+N+A+Distr	maskonâhk
in my many **book>s**	book	+N+Px1Sg+Pl	masinahikan+N+I+Px1Sg+Pl	nimasinahikana
my other little **tree** over there	tree	+N+Dim+Px1Sg+Obv	mîtos+N+A+Der/Dim+N+A+Px1Sg+Obv	nimîcosimisa

Appendix C. A detailed overview of the tags and their exponents for the English verb phrase generator (optional argument types in parentheses).[7] [8]

(Clause)	Tense Aspect[9]	Subject	(Direct Object)	(Indirect Object)	(Progressive)	(Repetitive)	(Negation)
Rel+: who; which Cnj+: as	Prs+: -(e)s Prt+: -(e)d Def+: will Fut+: is going to Int+: wants to Cond+:when Inf+: for s.o. to … Imm+: let s.o. … now Del+: let s.o. … later Prf1+: has done s.t. Prf2+: had done s.t.	0Sg+: it 1Sg+: I 2Sg+: you 3Sg+: he/she 1Du+: we both 2Du+: you both 3Du+: they both 1Distr+: each and every one of us 2Distr+: each and every one of you 3Distr+: each and every one of them: 0Pl+: they 1Pl+: we 21Pl+: you and we 2Pl+: you all 3Pl+: they 4Sg+: another 4Pl+: others 4Sg/Pl+: yet an/other(s) 5Sg/Pl+: yet others X+: someone XPl+: people	0SgO+: it 1SgO+: me 2SgO+: you 3SgO+: him/her 1DuO+: us both 2DuO+: you both 3DuO+: them both 1DistrO+: each and every one of us 2DistrO+: each and every one of you 3DistrO+: each and every one of them 0PlO+: them 1PlO+: us 21PlO+: you and us 2PlO+: you all 3PlO+: them 4Sg/PlO+: an/other(s) 5Sg/PlO+: yet an/other(s) XO+: someone XPlO+: people	1SgIO+: to me 2SgIO+: to you 3SgIO+: to him/her 1DuIO+: to us both 2DuIO+: to you both 3DuIO+: to them both 1PlIO+: to us 21PlIO+: to you and us 2PlIO+: to you all 3PlIO+: to them 4Sg/PlIO+: to an/other(s) 5Sg/PlIO+: to yet an/other(s) XIO+: to someone XPlIO+: to people	Prog+: be … -ing	Rept+: again and again; repeatedly	Neg+: not
(Modality) Obl2+: has to Nec2+: needs to Abl2+: is able to Perm2+: is allowed to Int2+: wants to Hab+: keeps on Init+: starts Fin+: finishes	**Auxiliary** Obl+: must Nec+: needs to Abl+: can Perm+: may Int+: wants to Poss+: could Rec+: should Pred+: would						

[7] The features used for English verb phrase generation are a superset of the features available for English verb phrase analysis. Furthermore, the tags for English verb phrase analysis are similar to the ones for English verb phrase generation, with the analysis tags preceded by a plus sign (e.g. `+1Sg` or `+Prt`) and all the tags following the English core lexical content resulting from the analysis (e.g. "I slept well" → `sleep well +Prt+1Sg`), whereas the generation tags are followed by a plus sign and all the tags precede the English core sentence frame that is to be manipulated (e.g. `Prt+1Sg+s/he sleeps well` → "I slept well").

[8] Example with maximal types of generation tags: `Abl2+Fut+1Sg+2SgO+3PlIO+Prog+Rept+Neg+s/he transfer s.o. to s.b.` → "I am not going to be being able to transfer you to them again and again".

[9] Features for Tense/Aspect and Auxiliary (i.e. simple modality, which consists of auxiliary verbs which cannot be further inflected) are mutually exclusive, whereas features for Tense/Aspect or Auxiliary can be combined with features for Modality (which are periphrastic constructions where the first element is treated as the finite verb that can be fully inflected).

Appendix D. A detailed overview of the tags and their exponents for the English noun phrase generator (optional argument types in parentheses).[10] [11]

Number/ Obviation	**(Diminutive)**	**(Possession)**
Sg+: a(n)	Dim+: little	Px1Sg+: my
Pl+: -(e)s		Px2Sg+: your
Obv+: an/other … -s		Px3Sg+: his/her
		Px1Pl+: our
Locative/ Distributive		Px12Pl+: your and our
Loc+: in		Px2Pl+: your
Distr+: among		PxPx3Pl+: their
		Px4Sg/Pl+: another's/others'
		PxX+: someone's

[10] The features used for English noun phrase generation are (currently) equivalent to the features available for English noun phrase analysis. Furthermore, the tags for English noun phrase analysis are similar to the ones for English noun phrase generation, with the analysis tags preceded by a plus sign (e.g. `+Pl` or `+Px1Sg`) and all the tags following the English core lexical content resulting from the analysis (e.g. ”my little black books” → `black book +N+Dim+Px1Sg+Pl`), whereas the generation tags are followed by a plus sign and all the tags precede the English core phrase frame that is to be manipulated (e.g. `Pl+Dim+Px1Sg+ a black book` → ”my little black books”). Importantly, in noun phrase generation a space character is currently required after the tags, before the phrase to be manipulated. Note also that the order of tags differs between noun phrase analysis, i.e. `(Diminutive) (Possession) Number/Locative`, and noun phrase generation, i.e. `Number/Locative (Diminutive) (Possession)`.

[11] Examples with maximal combinations of generation tags: `Obv+Dim+Px12Pl+ a black book` → “our and our other little black book(s)”; `Loc+Dim+Px4Sg/Pl+ a black book` → “in another's/others' little black book”.

Appendix E. Screenshots of *itwêwina*, with the English phrase analysis and generation integrated, exemplified with searches with “I helped you” and its approximate Cree match *kikîwîcihitin* (with inflectional morpheme boundaries marked with middle-dots).