Modeling the Noun Morphology of Plains Cree

Conor Snoek[1], *Dorothy Thunder*[1], *Kaidi Lõo*[1], *Antti Arppe*[1], *Jordan Lachler*[1]
Sjur Moshagen[2], *Trond Trosterud*[2]

[1] University of Alberta, Canada
[2] University of Tromsø, Norway

snoek@ualberta.ca, dthunder@ualberta.ca, kloo@ualberta.ca,
arppe@ualberta.ca, lachler@ualberta.ca,
sjur.n.moshagen@uit.no, trond.trosterud@uit.no

Abstract

This paper presents aspects of a computational model of the morphology of Plains Cree based on the technology of finite state transducers (FST). The paper focuses in particular on the modeling of nominal morphology. Plains Cree is a polysynthetic language whose nominal morphology relies on prefixes, suffixes and circumfixes. The model of Plains Cree morphology is capable of handling these complex affixation patterns and the morphophonological alternations that they engender. Plains Cree is an endangered Algonquian language spoken in numerous communities across Canada. The language has no agreed upon standard orthography, and exhibits widespread variation. We describe problems encountered and solutions found, while contextualizing the endeavor in the description, documentation and revitalization of First Nations Languages in Canada.

1 Introduction

The Department of Linguistics at the University of Alberta has a long tradition of working with First Nations communities in Alberta and beyond. Recently a collaboration has begun with Giellatekno, a research institute at the University of Tromsø, which has specialized in creating language technologies, particularly for the indigenous Saami languages of Scandinavia, but also for other languages that have received less attention from the computational linguistic mainstream. This collaboration is currently focusing on developing computational tools for promoting and supporting literacy, language learning and language teaching. Plains Cree is a morphologically complex language, especially with regard to nouns and verbs.

While we are working to develop a complete finite-state model of Plains Cree morphology, we focus on nominal morphology in this paper.

In the first section we briefly describe Plains Cree nominal morphology and give some background on the language. This is followed by details on the model and its implementation. Finally, we discuss the particular situation of developing tools for a language that lacks a formal, agreed-upon standard and the challenges that this presents. We conclude with some comments on the benefits of this technology to language revitalization efforts.

2 Background

2.1 Plains Cree

Plains Cree or *nêhiyawêwin* is an Algonquian language spoken across the Prairie Provinces in what today is Canada. It forms part of the Cree-Montagnais-Naskapi dialect continuum that stretches from Labrador to British Columbia. Estimates as to the number of speakers of Plains Cree vary a lot and the exact number is not known, from a high of just over 83,000 (Statistics Canada 2011, for Cree without differentiating for Cree dialects) to as low as 160 (Ethnologue 2013). Wolfart (1973) estimated there to be about 20,000 native speakers, but some recent figures are more conservative.

Regardless of the exact number of speakers, there is general agreement that the language is under threat of extinction. In many, if not most, communities where Cree is spoken, children are learning English as a first language, and encounter Cree only in the language classroom. However, vigorous revitalization efforts are underway and Cree is regarded as one of the Canadian First Nations languages with the best chances to prosper (Cook and Flynn, 2008).

As a polysynthetic language (Wolvengrey,

2011, 35), Plains Cree exhibits substantial morphological complexity. Nouns come in two gender classes: animate and inanimate. Each of these classes is associated with distinct morphological patterns. Both animate and inanimate nouns carry inflectional morphology expressing the grammatical categories of number and locativity. The number suffixes for animate and inanimate nouns are different, the plural being marked by *-ak* in animates and *-a* in inanimates. Locativity is marked by a suffix taking the form *-ihk* (with a number of allomorphs). The locative suffix cannot co-occur with suffixes marking number or obviation, but does occur in conjunction with possessive affixes. Obviation is a grammatical category marked on animate nouns that indicates relative position on the animacy hierarchy, when there are two third person participants in the same clause. Obviation is expressed through the suffix *-a*, which forms a mutually exclusive paradigmatic structure with the locative and number prefixes.

The possessor of a noun in Plains Cree is expressed through affixes attached to the noun stem. These affixes mark person and number of the possessor by means of a paradigmatic inflectional pattern that includes both prefixes and suffixes. Since matching prefixes and suffixes need to co-occur with the noun when it is possessed, it is possible to treat such prefix-suffix pairings as circumfixes expressing a single person-number meaning. The noun *maskisin* in (1) below[1] is marked for third person plural possessors as well as being plural itself. The inanimate gender class is recognizable in the plural suffix *-a*, which would be *-ak* in the case of an animate noun.

(1)
omaskisiniwâwa
o-maskisin-iwâw-a
3PL.POSS-shoe-3PL.POSS-PL.IN
'their shoes'

Nouns also occur with derivational morphology in the form of diminutive and augmentative suffixes. The diminutive suffix is productive and forms taking the diminutive suffix can occur with all the inflectional morphology described above.

(2)
omaskisinisiwâwa
o-maskisin-is-iwâw-a
3PL.POSS-shoe-DIM-3PL.POSS-PL.IN
'their little shoes'

The particular form of the diminutive, however, varies considerably. For example, the most common form of the suffix is *-is*. The suffix triggers morphophonemic changes in the stem. For example, the '*t*' in *oskâtâskw-* 'carrot' changes to '*c*' (the alveolar affricate [ts]) when the diminutive suffix is present resulting in the form *oskâcâskos*. Since the form *oskâtâskw-* is a *-w* final form a further phonological change occurs, namely the initial vowel in the suffix changes from *i > o*.

To sum up, Plains Cree nominal morphology allows the following productive pattern types:

(3)
stem+NUM
stem+OBV
stem+LOC
stem+DIM+NUM
stem+DIM+OBV
stem+DIM+LOC
POSS+*stem*+POSS+NUM
POSS+*stem*+DIM+POSS+NUM
POSS+*stem*+DIM+POSS+OBV
POSS+*stem*+POSS+LOC
POSS+*stem*+DIM+POSS+LOC

Plains Cree can be written both with the Roman alphabet and with a Syllabary. Theoretically there is a one-to-one match between the two. However, a number of factors complicate this relationship. Differing punctuation conventions, such as capitalization, and the treatment of loanwords make conversion from one writing system to another anything but a trivial matter. Orthography presents a general problem for the development of computer-based tools, because unlike nationally standardized languages, orthographic conventions can vary considerably from community to community, even from one user to another. Certain authors have argued for the adoption of orthographic standards for Plains Cree (Okimâsis and Wolvengrey, 2008), but simply is no centralized institution to enforce orthographic or other standardization. This means

---

[1] The following abbreviations are used POSS = possessive prefix/suffix; LOC = locative suffix; OBV = obviative suffix; DIM = diminutive suffix; NUM = number marking suffix; IN = inanimate; PL = plural.

that the wealth of varying forms and dialectal diversity of the language are apparent in each individual community. This situation poses specific challenges to the project of developing language tools that are more seldom encountered when making spell-checkers and language learning tools for more standardized languages.

Similar situations have been encountered in work on the Saami languages of Scandinavia (Johnson, 2013). Following their work, we include dialectical variants in the model, but mark them with specific tags. This permits a tool such as a spell-checker to be configured to accept and output a subset of the total possible forms in the morphological model. An example here is the distribution of the locative suffix described in more detail in section 4. There is a disparity between communities regarding the acceptability of the occurrence of the suffix with certain nouns. The suffix can be marked with a tag in the FST-model. This tag can then be used to block the acceptance or generation of this particular form. The key notion here is that language learning and teaching tools are built on the basis of the general FST model. For Plains Cree there is one inclusive model, encompassing as much dialectical variation as possible. From this, individual tools are created, e.g. spell-checkers, that selects an appropriate subset of the dialectically marked forms. A community can therefore have their own spell-checker, specific to their own preferences. It is also possible to allow for "spelling relaxations" (Johnson, 2013, 67) at the level of user input, meaning that variant forms will be recognized, but constraining the output to a selection of forms deemed appropriate for a given community. Hence, the spell-checker used in one particular community could accept certain noun-locative combinations. At the same time, other tools, such as paradigm learning applications, could block this particular noun-locative combination from being generated: certain forms are understood, but not taught by the model. In general, the variation is not difficult to deal with in terms of the model itself, rather it represents a difficulty in the availability of accurate descriptions, since their specifics must be known and understood to be successfully included in the model.

This method could, in principle, be used to extend the Plains Cree FST-model to closely related Algonquian languages. However, rather than creating a proliferation of dialectical tags, it is easier to reproduce the architecture of the model and use it to create a new model for the related language. This allows the preservation of formal structures that follow essentially the same pattern, such as possessive inflection for example, while replacing the actual surface forms with those of the target language.

2.2 Previous computational modeling of Algonquian languages

Previous work on Algonquian languages that has taken a computational approach is not extensive. Hewson (1993) compiled a dictionary of Proto-Algonquian terms generated through an algorithm. His data were drawn from fieldwork carried out by Leonard Bloomfield. Kondrak (2002) applied algorithms for cognate identification to Algonquian data with considerable success. **?**) worked on a sizable corpus of Cree data and developed tools for data management and analysis in PL/I. Junker (2008) have written on the difficulties of creating search engine tools for East Cree and describe challenges similar to the ones we have encountered with regard to dialectal variation and the absence of agreed on standard orthographies and other widespread conventions.

In general, computational approaches to Algonquian, and other Indigenous North American languages, have been hampered by the fact that in many cases large bodies of data to develop and test methods on are just not available. Even for Plains Cree, which is relatively widely spoken, and relatively well documented, the available descriptions are still lacking in many places. As a result, fieldwork must be undertaken in order to establish patterns that can be modeled in the formalism necessary for the finite state transducer (FST) to work, a point that will be expanded on below.

3 Modeling Plains Cree morphology

The finite state transducer technology that forms the backbone of our morphological model, and consequently of all the language applications we are currently developing, is based historically on work on computational modeling of natural languages known as two-level morphology (TWOL) by Koskenniemi (1983). His ideas were further developed by Beesley and Karttunen (2003). Their framework offers two basic formalisms with which to encode linguistic data, *lexc* and *twolc*.

The Lexicon Compiler, or *lexc*, is "a high-level declarative language and associated compiler" (Beesley and Karttunen, 2003, 203) used for encoding stem forms and basic concatenative morphology. The source files are structured in terms of a sequence of continuation lexica. Beginning with an inventory of stems the continuation lexica form states along a path, adding surface morphological forms and underlying analytic structure at each stage. A colon (:) separates underlying and surface forms. Example (4) demonstrates paths through just three continuation lexica for the animate nouns *apiscacihkos* 'antelope' and *apisimôsos* 'deer'. By convention, the names of continuation lexica are given in upper case. Stems and affixes represent actual word forms, and are thus given in lower case. The '+' sign indicates a morphological tag.

(4)
```
LEXICON ANSTEMLIST
apiscacihkos ANDECL ;
apisimôsos ANDECL ;
LEXICON ANDECL
< +N:0 +AN:0 +Sg:0 @U.noun.abs@ # > ;
< +N:0 +AN:0 @U.noun.abs@ OBVIATIVE > ;
LEXICON OBVIATIVE
< +Obv:a # > ;
```

Both forms are directed to the continuation lexicon here named ANDECL which provides some morphological tagging in the form of +N to mark the word as a noun and +AN to denote the gender class 'animate'. Each of the two nouns has the possibility of passing through the continuation lexicon ANDECL as an 'absolutive' noun – as indicated by the tag `@U.noun.abs@` (a *flag diacritic*, as will be explained below). The colons in the code indicate a distinction between upper and lower levels of the transducer. The upper form to the left of the colon is a string containing the the lemma as well as a number of tags that contain information about grammatical properties. For the word form *apiscacihkos*, the analysis once it has passed through the ANDECL continuation lexicon is `apiscacihkos+N+AN+Sg`.

The surface forms *apiscacihkos* and *apisimôsos* are well-formed strings of Plains Cree, following the Standard Roman Orthography. Hence, the path can terminate here as indicated by the hash mark. The other path, also open to both forms

since they pass through the same continuation lexicon, leads to a further continuation lexicon named OBVIATIVE. This rather small lexicon adds a final *-a* suffix and the tag `+Obv` indicating that the form is inflected for the grammatical category of obviation. Since no number suffixes can occur in this form the path does not add a `+Sg` or `+Pl` tag to the underlying form.

(5)
```
apiscacihkos+N+AN+Obv
```
apiscacihkosa
'antelope'

These circumfixes were modeled using Flag Diacritics, which are an "extension of the finite state implementation, providing feature-setting and feature-unification operations" (Beesley and Karttunen, 2003, 339). Flag diacritics make it possible for the transducer to *remember* earlier states. The transducer may travel all paths through the prefixes via thousands of stems to all the suffixes, but the flag diacritics ensure that only strings with prefixes and suffixes belonging to the same person-number value are generated. In our solution for nouns, the continuation lexica allow all combinations of possession suffixes and prefixes, but the flag diacritics serve to filter out all undesired combinations. For example, in the noun *omaskisiniwâwa* from (1) above, the third person prefix *o-* and the suffix marking both person and number *-iwâw* are annotated in the *lexc* file with identical flag diacritics, so that they will always occur together.

Plains Cree has some very regular and predictable morphophonological alternations that can be modeled successfully in the finite state transducer framework. The formalism used here is not *lexc* as in the listing of stems and the concatenative morphology, but an additional formalism called the two-level compiler or *twolc* that is well suited to this task. The *twolc* formalism was developed by Lauri Karttunen, Todd Yampol, Kenneth R. Beesley and Ronald M. Kaplan based on ideas set forth in Koskenniemi (1983).

(6)
acâwewikamikosis
atâwewikamikw-isis
store-DIM
'little store'

In example (6) above, the form *atâwewikamikw* 'store' is modified by the derivational suffix *-isis* marking the diminutive form. This derivation is highly productive in Plains Cree. The underlying form of the suffix is *-isis* but in conjunction with a stem-final *-w*, the initial vowel of the suffix changes to *-o*. This morphophonemic alternation can be written in $twolc$ much like a phonological rule:

(7)
```
i:o <=> w: %>:0 _ s: +Dim ;
```

The sign %> is used to mark a suffix boundary, which, along with the +Dim tag, ensures that it is the first vowel of the suffix that undergoes substitution. Thus the context is given by the occurrence of a *-w* before the suffix boundary, i.e. stem finally. An additional complication here is that the presence of the diminutive suffix in a form again triggers a phonological change in the stem by which all *t*'s change to *c*'s (phonetically [ts]). In $twolc$ the rules dictating morphophonological alternations apply in parallel, avoiding possible problems caused by sequential rule interactions. The noun completes the path through the continuation lexica and is passed to $twolc$ as `atâwewikamikwisis`. There it undergoes two morphophonological changes giving the correct surface form *acâwewikamikosis*.

Twolc is a powerful mechanism for dealing with regular alternations. Reliance on $twolc$ can reduce the number of continuation lexica and hence complexity of the morphology modeling carried out in $lexc$. The downside of using large numbers of $twolc$ rules is the increasing complexity of rule interactions. We have found that decisions about which strategy to pursue in the modeling of a particular morphological pattern must frequently be made on a case by case basis. For example, in modeling the interesting case of the form *atimw-* 'dog' several strategies needed to be employed. The form triggers a vowel change $i > o$ in conjunction with the diminutive suffix *-isis* resulting in *-osis*, a change falling under a rule described in (8) above. A further change here is that the *t* in *atimw-* 'dog' changes to *c* when the diminutive suffix is present resulting in the surface form *acimosis*. Both these forms can be handled by $twolc$ rules such as the one exemplified in (8) above.

However, *atimw-* also undergoes changes in the stem vowel when the noun is marked for a possessor so that $a > i$ and $i > ê$. In the first person, the possessive prefix takes the form *ni-* leading to a sequence of two vowels arising from the prefix final *-i-* and stem initial *-i-*, which is not permitted in Plains Cree. This situation is handled by a general rule deleting the first vowel in preference for the latter. However, a set of $twolc$ rules would be required to change the stem vowels – a set that would be specific to this particular word only. The full set of two level rules are accessible online[2].

Since the addition of further rules poses the risk of rule conflicts in an increasingly complex $twolc$ code, the stem vowel changes are handled in $lexc$ instead. There are currently over 40 continuation lexica in the model of nominal morphology alone.

(8)
```
LEXICON IRREGULARANIMATESTEMS
atim IRREGULARINFLECTION-1 ;
atim:têm IRREGULARINFLECTION-2 ;
```

The continuation lexicon contains two versions of the form *atim* with two different paths leading to further inflectional suffixes. In the second instance of *atim*, writing the base form to the left of the colon and the suppletive stem to the right ensures both that the form *-têm* surfaces correctly. In the analysis the base form *atim* can still be recovered. The forms are sent to differing continuation lexica, since only the suppletive forms occurs within the paradigm of possessive prefixes. The word meaning 'my little dog' is given as an example in (10) below.

(9)
nicêmisis
ni-atimw-isis
1SG.POSS-dog-DIM
'my little dog'

The suppletive form also does not carry an underlying *-w* and hence no longer triggers the vowel change in the diminutive suffix. With this solution we can handle the regular and more straightforward morphophonological alternations in $twolc$, while avoiding undue complexity by modeling the suppletive forms in $lexc$.

Finally, we have adopted a system of using special tags to denote dialectal variants that are not equally acceptable in different communities. The seemingly high level of variation found in Plains Cree can be related to several reasons described in more detail in the next section. The variation is dealt with in the morphological model with a tagging strategy that marks dialectal forms. This tagging allows for the systems based on the morphological model to behave in accordance with the wishes of the user or community of users. In the setting of a particular teaching institution, for instance, only a certain subset of the variants encoded in the morphological model might be deemed acceptable. Our model permits this community to adjust the applications they are employing, e.g. a spell-checker, so that their community-specific forms are accepted as correct.

The stems are accessible online[3], and may be analysed and generated at the webpage for Plains Cree grammar tools[4].

4 The necessity for fieldwork in modeling Plains Cree

We began working on the morphological model of Plains Cree by examining published sources, such as *Plains Cree: A grammatical study* (Wolfart, 1973) and *Cree: Language of the Plains* (Okimâsis, 2004). Okimâsis' work is clearly structured and contains a wealth of information. Nevertheless, the level of explicitness required to capture the nature of a language in enough detail for applications such as, for example, spell-checkers is beyond the scope of her work. This is to say that in formalizing Okimâsis' description we needed to generalize grammatical patterns that were not always explicitly spelled out in her work in every detail. It should be apparent here that a number of factors come in to play here that make working on Plains Cree quite a different undertaking from working on a European language with a long history of research in the Western academic tradition. While official national European languages such as German, Finnish or Estonian can look on scholarly work dating back some centuries, and are supported by work from a community of specialists numbering hundreds of people, work on Plains Cree (and other languages in

similar situations) is being carried out by what is at best a handful of people. While Cree language specialists form a professional body of researchers with a proud tradition, they are faced with the enormous task of documenting a language spoken in many small communities spread over a huge geographical area. In addition, many of those specialists are also involved with language revitalization and language teaching, with the result that less time can be devoted to language description, scholarship and the pursuit of larger projects such as the development of corpora. While such projects are under development in many areas, the demands placed on individual researchers and activists has resulted in an overall scarcity of resources. While compared to other Indigenous languages spoken in Canada, Plains Cree is relatively well documented, many of the resources that would be desirable assets for the development of a finite state model are not available. As a result, we have carried out fieldwork to further make explicit the full inflectional paradigm of nouns in Plains Cree.

There is considerable variation among speakers and specialists regarding the acceptability of certain inflectional possibilities. For example, in the case of one animate noun *atim* 'dog' it seems formally reasonable to allow its combination with the locative suffix *-ohk* rendering *atimohk*. This combination of stem and affix was considered impossible or at least implausible by some of our native speaker consultants. However, the form itself does occur, albeit in the guise of a place name for a lake island in northern Saskatchewan named *atim* 'dog'. Therefore the form *atimohk* 'on the dog' with locative suffix attached can occur in this very specific and geographically bounded context[5]. The way of coping with this is to lexicalize *atimohk* as locative of the island *Atim*, and to keep the noun *atim* outside the set of nouns getting regular locatives.

Further inquiry into this matter revealed that some speakers see the locative suffix as potentially occurring quite widely, while others are more restrictive (Arok Wolvengrey – p.c.). Here again there is a problem of scale: individual speakers of any language have only a partial experience of the possible extent of the language. In the modeling of the morphology for the purposes of such technologies as spell-checkers, for example, the expe-

---

rience of any potential speaker must be taken into account. While the information that this particular form is rare or semantically not well-formed is valuable, retaining the form is important, if the model is to cover the range of potential usage patterns of all Plains Cree speakers. Ideally, if the written use of the language is supported by the tools that can be developed based on our morphological model, that would lead to a gradually increasing electronic corpus of texts, providing frequency information on both the stems and morphological forms.

We have developed a workflow in which we construct the maximal paradigms that are theoretically possible and then submit them to intense native speaker scrutiny. Only once native speakers and specialists have approved the forms do they become part of the actual model. The paradigms are chosen so as to provide the coverage of the entire span of morphologically possible forms as well as all morphophonemic alternations. As such they present a maximal testbed for the patterns encoded in the formalism. Each paradigm consists of about sixty inflected forms.

Overall, a careful balance must be struck between directly explicit speaker/specialist input and theoretically possible forms. We aim to achieve this balance by taking a threefold approach: First, by careful consultation with speakers and specialists; second, by building a corpus[6] which can serve as a testing ground for the morphological analyzer and as a source of data, and third by working closely with communities willing to test the model and provide feedback.

5 Applications in language teaching and revitalization

The development of an explicit model of the morphology of Plains Cree as outlined above is of benefit not just to researchers but also those involved in teaching and revitalizing the language within their home communities. Using the general technological infrastructure developed by the researchers at Giellatekno, we are able to take the FST model of Plains Cree morphology and use it to create in one go a variety of language tools including a spellchecker, a morphological analyzer and a paradigm generator, which can be integrated

as modules within general software applications such as a word-processor, an electronic dictionary or a intelligent computer-aided language learning (CALL) application. Each of these tools can assist fluent speakers, as well as new learners, in their use of Plains Cree as a written language.

The spellchecking functionality within a word-processor will be a valuable tool for the small-but-growing number of Plains Cree language professionals who are engaged in the development of teaching and literary resources for the language. It will allow for greater accuracy and consistency in spelling, as well as faster production of materials. Because dialectal variation is being encoded directly into the FST model, the spellchecker can be configured so that writers from all communities and dialects can use this tool, without worry that the technology is covertly imposing particular orthographic standards which the communities have not all agreed upon.

The morphological analysis functionality built from the FST model and integrated within e.g. a web-based electronic dictionary will allow readers to highlight Plains Cree text in a document or web-page to perform a lookup of words in any inflected form, and not only with the citation (base) form. This will enable readers to more easily read Plains Cree documents with unfamiliar words without needing to stop to repeatedly consult paper dictionaries and grammars. While this does not obviate the need for printed resources in learning and teaching of the language, such added functionality can greatly increase the pace at which texts are read through by language learners. This is not inconsequential as it can slow down considerably the onset of weariness brought on by needing to interrupt the reading process to consult reference materials, and hence maintain the motivation for language learning.

The paradigm generation functionality within e.g. an electronic dictionary allows users to select a word and receive the full, or alternatively a smaller but representative, inflected paradigm of that word. This will be of direct benefit to instructors developing materials to teach the complex morphology of the Plains Cree, as well as their students.

We are working in collaboration with Plains Cree communities in the development and piloting of these tools, to ensure their accuracy and their usefulness for teachers, developers, learners and

---

[6]As noted above, a tool like a spell-checker promotes literacy and hence contributes naturally to the increase in textual materials. Until that begins to happen, however, we are collecting texts through recording and transcription.

other community members. The full range of uses that these tools will be put to will only become apparent over time, but we expect that they will have a positive impact for community language maintenance by supporting the continued development Plains Cree literacy.

6 Conclusion

We have found the technology of Finite State Transducers so useful in developing language applications for Plains Cree because it permits us to integrate native speaker competence and specialized linguistic understanding of grammatical structures into the model directly.

At present the analyzer contains 72 nominal lexemes, carefully chosen to cover all morphological and morphophonological aspects of the Plains Cree nominal system. Once the morphological modeling of this core set of nouns has been finalized, scaling up the lexicon will be a trivial task, as all lexicographic resources classify their stem in the same way as is done in the morphological transducer.

We have described our method of working with native speaker specialists and how their insights are reflected in the design of the model. This interaction also allows enough possibilities for interactions with language teachers, learners and activists so that we make our work truly useful to the effort of preserving and revitalizing the precious cultural heritage that is Plains Cree. We are aware of the limits of tools that relate primarily to the written forms for languages that have rich oral histories and cultures, but feel that writing and reading Plains Cree will play an ever growing role in the future of this language.

This work makes practical contributions to linguistic research on Plains Cree. On the one hand, creating the model required the formalization of many aspects of Plains Cree morphology which had not previously been spelled out in full detail, i.e. it makes explicit what is known, or not known, about Plains Cree morphology, and thus allows us to extend the description of Plains Cree morphology accordingly. On the other, the morphological analyses can aid in future linguistic discovery especially when used in conjunction with corpora.

In the future, we will continue to expand the morphological model both in its grammatical coverage and in the size of the lexical resources which go into it. In regard to the latter, we are working with Cree-speaking communities in Alberta to expand on existing dictionaries and develop collections of recordings. The development of this morphological model has led us to carry out fieldwork on Plains Cree and to actively engage with Cree-speaking communities. We have worked hard to bridge the unfortunate gap that sometimes forms between the linguistic work being carried within academia and the needs of communities that are active in language documentation and revitalization. We look forward to further fruitful cooperation between activists, educators and researchers.

References

Kenneth R. Beesley and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Publications, Stanford

(CA).

Eung-Do Cook and Darin Flynn. 2008. Aboriginal languages of Canada. In: O'Grady, William and John Archibald (eds.) *Contemporary Linguistic Analysis*. Pearson, Toronto (ON).

John Hewson. 1993. *A computer-generated dictionary of Proto-Algonquian*, Canadian Museum of Civilization and Canadian Ethnology Service, Ottawa (ON).

Ryan Johnson, Lene Antonsen and Trond Trosterud. 2013. Using Finite State Transducers for Making Efficient Reading Comprehension Dictionaries. In Stephan Oepen & Kristin Hagen & Janne Bondi Johannessen (eds.), *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODAL-IDA 2013)*, 378-411. Linköping Electronic Conference Proceedings No. 85.

Marie-Odile Junker and Terry Stewart. 2008. Building Search Engines for Algonquian Languages. In Karl S. Hele & Regna Darnell (eds.), *Papers of the 39th Algonquian Conference*, 59-71. University of Western Ontario Press, London (ON).

Grzegorz Kondrak. 2002. *Algorithms for Language Reconstruction*, Department of Computer Science, University of Toronto.

Kimmo Koskenniemi. 1983. *Two-level Morphology: A General Computational Model for Word-Form Recognition and Production*, Publication No. 11. Department of General Linguistics, University of Helsinki.

Jean Okimâsis. 2004. *Cree: Language of the Plains*, Volume 13 of University of Regina publications. University of Regina Press, Regina (SK).

Jean Okimâsis and Arok Wolvengrey. 2008. *How to Spell it in Cree*. miywâsin ink, Regina (SK).

H. Christoph Wolfart. 1973. Plains Cree: A grammatical study, *Transactions of the American Philosophical Society No. 5*.

H. Christoph Wolfart and Francis Pardo 1973. *Computer-assisted linguistic analysis*, University of Manitoba Anthropology Papers No. 6. Department of Anthropology, University of Manitoba.

Arok E. Wolvengrey. 2011. *Semantic and pragmatic functions in Plains Cree syntax*, LOT, Utrecht (NL).