



Lexical Subtypes in a Plains Cree Corpus

**Atticus G. Harrigan
Katherine Schmirler
Antti Arppe**
October 13-16, 2016

Introduction

Background

Our corpus

What is core?

Method

Results

Conclusion

References

Introduction

Who are we?

- **Alberta Language Technology Laboratory (ALT Lab)**
- **International group of collaborators from within Canada (FNUC) and beyond (Giellatekno at UiT)**
- **Team of theoretical, descriptive, corpus, and computational linguists**

What do we do?

- **Focus on the creation of linguistic tools, especially for understudied languages**
- **Project working on Algonquian, Siouan, Athabaskan, and Haida languages**
- **Tools take the form of smart online dictionaries, speech recognition, speech synthesis, spell checkers, corpus construction, etc.**

Our Corpus

- **Wolfart-Ahenakew Corpus**
- **~100, 000 words**
- **Multiple speakers; Elicited by Dr. Ahenakew**
- **Fully Cree discourse, often conversational**
- **Made available to our research team by Dr. Wolfart**

What is Core?

Core (based on Karlsson 1985)

- A subset of the full paradigm
- Only those forms that are *obligatory*
- All core forms are equal in value
- Core forms are in complementary distribution
- For Plains Cree, these forms are, minimally, person marking and conjunct marking preverbs (and theme for TA verbs)
- Unspecified actor marking is probably not core, as it is derivational and leads to a new set of inflections

Peripheral Forms

- Non-obligatory
- Non-complimentary in distribution
- Non-equal cells
- In Plains Cree, these are lexical preverbs and optional initials, medials, finals, etc.

What is Core?

Gradient Forms

- These forms have some aspect of core-ness and some aspect of peripheral-ness
- Grammatical preverbs like tense and conjunct marking are gradient as they are not necessary, but seem to occur often
- Future intentional tense (*ka-*) is likely most gradient, as *ka-kî-* ‘was going to’ is a valid string, showing non-complimentary distribution with the past tense *kî-*. Further, the future intentional morpheme can occur in the conjunct order

Method

Feature tagging

- Using our finite state transducer, we run each word of the corpus (as defined by white space on each side—we can do this due to the strict standardization of the corpus)
- Words are broken down into possible morpheme combinations and tagged for features such as person, tense, order, etc.
- When a word returns more than one result, the result with fewer morphemes is generally preferred
- Types (different word forms) and tokens (individual instances of a word form) are recorded

Count/feature retrieval

- Person is marked in the form of *1Sg*, and so we simple search for these forms and record the type and token counts.

NOTE: We present here only on the independent and conjunct orders. We have not studied the imperative (as this corpus lacked any examples for VAI or VTI forms) nor the future conditional (of which there was only one type with two tokens in the VTI: *wêwânêyihthkwâwi* ‘if they were worried’)

VII TYPES

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|------------|-----------|----------|----------|------------|
| 0 Sg | 126 | 31 | 6 | 4 | 167 |
| 0 Pl | 75 | 21 | 0 | 0 | 96 |
| 0' Sg | 7 | 5 | 0 | 0 | 12 |
| 0' Pl | 0 | 2 | 0 | 0 | 2 |
| TOTAL | 208 | 57 | 6 | 4 | 275 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|------------|------------|-----------|-----------|------------|
| 0 Sg | 236 | 64 | NA | 14 | 314 |
| 0 Pl | 75 | 35 | NA | 0 | 110 |
| 0' Sg | 51 | 9 | NA | 4 | 64 |
| 0' Pl | 20 | 4 | NA | 0 | 24 |
| TOTAL | 383 | 108 | NA | 18 | 509 |

VII TOKENS

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|------------|------------|----------|----------|------------|
| 0 Sg | 358 | 64 | 8 | 4 | 434 |
| 0 Pl | 36 | 45 | 0 | 0 | 81 |
| 0' Sg | 7 | 5 | 0 | 0 | 12 |
| 0' Pl | 0 | 2 | 0 | 0 | 2 |
| TOTAL | 401 | 116 | 8 | 4 | 529 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|------------|------------|-----------|-----------|-------------|
| 0 Sg | 640 | 131 | NA | 20 | 791 |
| 0 Pl | 102 | 44 | NA | 0 | 146 |
| 0' Sg | 76 | 9 | NA | 5 | 90 |
| 0' Pl | 22 | 4 | NA | 0 | 26 |
| TOTAL | 842 | 188 | NA | 25 | 1055 |

VAI TYPES

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|------------|------------|-----------|-----------|------------|
| 1 Sg | 162 | 64 | 17 | 12 | 255 |
| 2 Sg | 20 | 11 | 0 | 4 | 35 |
| 3 Sg | 156 | 110 | 13 | 7 | 286 |
| 1 Pl | 62 | 42 | 4 | 3 | 111 |
| 21 Pl | 12 | 3 | 2 | 2 | 19 |
| 2 Pl | 14 | 3 | 2 | 3 | 22 |
| 3 Pl | 108 | 88 | 6 | 0 | 202 |
| 3' | 17 | 5 | 0 | 0 | 22 |
| TOTAL | 551 | 326 | 44 | 31 | 952 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL |
|--------------|-------------|------------|-----------|-----------|-------------|
| 1 Sg | 349 | 118 | NA | 23 | 490 |
| 2 Sg | 70 | 14 | NA | 3 | 87 |
| 3 Sg | 549 | 271 | NA | 35 | 855 |
| 1 Pl | 248 | 150 | NA | 12 | 410 |
| 21 Pl | 61 | 4 | NA | 0 | 65 |
| 2 Pl | 91 | 12 | NA | 1 | 104 |
| 3 Pl | 421 | 206 | NA | 17 | 644 |
| 3' | 111 | 29 | NA | 2 | 142 |
| TOTAL | 1900 | 804 | NA | 93 | 2797 |

VAI TOKENS

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL |
|---------------|-------------|------------|-----------|-----------|-------------|
| 1 Sg | 290 | 73 | 22 | 19 | 404 |
| 2 Sg | 24 | 11 | 0 | 4 | 39 |
| 3 Sg | 924 | 196 | 16 | 8 | 1144 |
| 1 Pl | 82 | 58 | 4 | 3 | 147 |
| 21 Pl | 14 | 4 | 2 | 3 | 23 |
| 2 Pl | 16 | 4 | 2 | 5 | 27 |
| 3 Pl | 149 | 121 | 6 | 0 | 276 |
| 3' | 19 | 6 | 0 | 0 | 25 |
| TOTALS | 1518 | 473 | 52 | 42 | 2085 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL |
|---------------|-------------|-------------|-----------|------------|-------------|
| 1 Sg | 558 | 156 | NA | 36 | 750 |
| 2 Sg | 81 | 14 | NA | 3 | 98 |
| 3 Sg | 908 | 412 | NA | 39 | 1359 |
| 1 Pl | 353 | 196 | NA | 12 | 561 |
| 21 Pl | 78 | 4 | NA | 0 | 82 |
| 2 Pl | 111 | 12 | NA | 1 | 124 |
| 3 Pl | 604 | 296 | NA | 20 | 920 |
| 3' | 128 | 31 | NA | 3 | 162 |
| TOTALS | 2821 | 1121 | NA | 114 | 4056 |

VTI TYPES

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL | |
|--------------|-----|------------|------------|-----------|----------|------------|
| 1 Sg | | 81 | 24 | 2 | 4 | 111 |
| 2 Sg | | 18 | 3 | 3 | 1 | 25 |
| 3 Sg | | 33 | 32 | 3 | 1 | 69 |
| 1 Pl | | 15 | 20 | 1 | 1 | 37 |
| 21 Pl | | 6 | 1 | 0 | 0 | 7 |
| 2 Pl | | 10 | 1 | 1 | 0 | 12 |
| 3 Pl | | 28 | 31 | 4 | 1 | 64 |
| 3' | | 3 | 3 | 0 | 0 | 6 |
| TOTAL | | 194 | 115 | 14 | 8 | 331 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL | |
|--------------|-----|------------|------------|-----------|-----------|-------------|
| 1 Sg | | 154 | 69 | NA | 11 | 234 |
| 2 Sg | | 37 | 5 | NA | 8 | 50 |
| 3 Sg | | 219 | 123 | NA | 19 | 361 |
| 1 Pl | | 93 | 62 | NA | 4 | 159 |
| 21 Pl | | 19 | 6 | NA | 1 | 26 |
| 2 Pl | | 31 | 8 | NA | 1 | 40 |
| 3 Pl | | 153 | 114 | NA | 8 | 275 |
| 3' | | 18 | 6 | NA | 0 | 24 |
| TOTAL | | 724 | 387 | NA | 52 | 1163 |

VTI TOKENS

| IND | Prs | Pst | Fut Def | Fut Int | TOTAL | |
|--------------|-----|------------|------------|-----------|-----------|------------|
| 1 Sg | | 288 | 76 | 5 | 4 | 373 |
| 2 Sg | | 58 | 5 | 3 | 2 | 68 |
| 3 Sg | | 99 | 61 | 3 | 1 | 164 |
| 1 Pl | | 31 | 36 | 1 | 2 | 70 |
| 21 Pl | | 14 | 1 | 0 | 0 | 15 |
| 2 Pl | | 24 | 1 | 1 | 0 | 26 |
| 3 Pl | | 50 | 61 | 7 | 1 | 119 |
| 3' | | 5 | 2 | 0 | 0 | 7 |
| TOTAL | | 569 | 243 | 20 | 10 | 842 |

| CNJ | Prs | Pst | Fut Def | Fut Int | TOTAL | |
|--------------|-----|-------------|------------|-----------|-----------|-------------|
| 1 Sg | | 285 | 127 | NA | 22 | 434 |
| 2 Sg | | 49 | 5 | NA | 8 | 62 |
| 3 Sg | | 398 | 240 | NA | 27 | 665 |
| 1 Pl | | 204 | 85 | NA | 4 | 293 |
| 21 Pl | | 21 | 7 | NA | 1 | 29 |
| 2 Pl | | 37 | 9 | NA | 1 | 47 |
| 3 Pl | | 287 | 213 | NA | 8 | 508 |
| 3' | | 30 | 6 | NA | 0 | 36 |
| TOTAL | | 1311 | 692 | NA | 71 | 2074 |

Conclusion

Tense

- **Future intentional/definite tenses are fairly rare, especially in the VII and VTI classes. The VII class showed no future tense in the independent and few in the conjunct. This finding seems congruent with the idea that future tense, especially future intentional is not entirely core**
- **Present tense is much more prevalent than past tense in VAI, VII, and VTI classes**

Order

- **In all classes, the conjunct order is by far the more dominant order; in many cases, the conjunct was 2-3 times more common than the independent**

Person Marking

- **For VII: ~80% of tokens are in 0 Sg; the second most common person is 0 Pl. Obviative VII actors are very rare**
- **For VAI/VTI: 3 Sg is the most dominant case (except in VTI independent, where 1 Sg has more tokens), followed by 1 Sg / 3 Pl, followed by 1 Pl. Second person forms, even 21 Pl were very rare, with 21 Pl being the least numerous of all person markings**

Wolfart-Ahenakew Corpus:

- Ahenakew A. (2000). âh-âyîtaŵ isi ê-kî-kiskêyihthakik maskihkiy/They knew both sides of medicine: Cree tales of curing and cursing told by Alice Ahenakew, edited by H.C. Wolfart and F. Ahenakew. Publications of the Algonquian Text Society Winnipeg: University of Manitoba Press
- Bear G., Fraser M., Calliou I., Wells M., Lafond A., Longneck R. (1992). kôhkominawak otiâcimowiniwâwa/Our Grandmothers' Lives: As Told In Their Own Words, edited by F. Ahenakew and H.C. Wolfart, vol 3. University of Regina Press
- Kâ-Nîpitêhtêw J. (1998). Counselling Speeches of Jim Kâ-Nîpitêhtêw, edited by F. Ahenakew and H.C. Wolfart. University of Manitoba Press
- Masuskapoe C. (2010). piko kîkway ê-nakacihtât: kêkêk otâcimowina ê-nêhiyawastêki, edited by H.C. Wolfart and F. Ahenakew. Algonquian and Iroquoian Linguistics
- Minde E. (1997). kwayask ê-kî-pê-kiskinowâpahtihicik/Their Example Showed Me the Way, edited by F. Ahenakew and H.C. Wolfart. Edmonton: University of Alberta Press
- Vandall P., Douquette J. (1987). wâskahikaniwiyiniw-âcimowina/Stories of the House People, edited by F. Ahenakew. University of Manitoba Press
- Whitecalf S. (1993). kinêhiyâwiwininaw nêhiyawêwin/The Cree language is our identity: The La Ronge lectures of Sarah Whitecalf, edited and translated by H.C. Wolfart and F. Ahenakew. Publications of the Algonquian Text Society / Collection de la Société d'édition des textes algonquiennes Winnipeg: University of Manitoba Press

Bibliography:

- Karlsson, F. (1985). Paradigms and word forms. *Studia gramatyczne*, 7, 135-154.
- Karlsson, F. (1986). Frequency considerations in morphology. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39(1), 19-28.
- Okimasis, J. (2004). *Cree, Language of the Plains*. Regina: University of Regina Press.
- Schmirler, K. & Harrigan, A. 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.
- 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. Vol. 17: Languages* (pp. 390-439). Washington: Smithsonian Institute.
- Wolvengrey, A. (2001). *nēhiyawēwin: itwēwina / Cree: Words*. Regina: University of Regina Press.

Thank you!
Questions?

galvin@ualberta.ca
schmirle@ualberta.ca
arppe@ualberta.ca

Alberta Language Technology Lab: <http://atllab.artsrn.ualberta.ca/>