UNIVERSITY OF ALBERTA

A quantitative look at Plains Cree text types:

âtayôhkêwina vs. *âcimowina* in Bloomfield's texts and *âcimisowina* vs. *kakêskihkêmowina* in the Ahenakew-Wolfart corpus

or, How (Not) To Do Textual Analysis

Katherine Schmirler & Antti Arppe
University of Alberta

52nd Algonquian Conference
October 23, 2020

Plains Cree text types

- Two broad categories:

- *âtayôhkêwina* (sacred stories)
- *âcimowina* (other stories)
  - *âcimisowina* (stories about oneself)
  - *wayiwatâcimowina* (funny stories)
  - *kayâs-âcimowina* (old-time stories)
  - *kakêskihkêmowina* (counselling speeches)

- *pîkiskwêwina* (dialogues)

Ahenakew-Wolfart

Bear et al. (1998): *âcimisowina*

Kâ-Nîpihtêhtêw (1998) & Whitecalf (1993): *kakêskihkêmowina*

	âcimisowina	**kakêskihkêmowina**
Tokens	20007	15930
Types	4118	3390
Cree tokens	**11051**	**10231**
Cree types	**3615**	**3084**

Bloomfield

Plains Cree Texts (1934): *âcimowina & âtayôhkêwina*

Sacred Stories of the Sweetgrass Cree (1930): *âtayôhkêwina*

	âcimowina	**âtayôhkêwina**
Tokens	37301	65660
Types	7053	10615
Cree tokens	**26736**	**45736**
Cree types	**7039**	**10601**

Register analysis

- Following Biber and colleagues (e.g. Biber, 1991; Biber et al., 1998, 2002; Biber & Conrad, 2019)

1) Situational context: what is the purpose of a text? Who is speaking, to whom, for what reason?
2) Linguistic features: what features demonstrate distributional differences between the two texts?
3) Bring it all together: what links the features and the situational context?
4) Rinse and repeat: what does the analysis tell us about the situational context, etc.?

A bottom-up approach

- Exploring the relative frequencies of morphosyntactic features (Harrigan et al., 2017; Schmirler et al., 2018; Arppe et al., 2020):
  - Morphological model output
    - How many verbs are there? How many of each transitivity class? Order? Person? How many are quotative verbs?
    - How many nouns? How many animate, inanimate, dependent, possessed?
    - How many pronouns? Of which types?
  - Syntactic parser output
    - How many clauses contain verbs? How many verbs have overt actors/goals?
- But to keep things accessible for a short presentation: we'll just talk about the ones that stand out

UNIVERSITY OF ALBERTA

âcimisowina & kakêskihkêmowina

in the Ahenakew-Wolfart corpus

Verbs

	âcimisowina		**kakêskihkêmowina**	
	Tokens	Types	Tokens	Types
Verbs	3168 (28.7%)	2252 (62.3%)	3084 (30.1%)	2238 (69.0%)
	Of verbs…		**Of verbs…**	
VAI	**1505 (47.5%)**	955 (42.4%)	1006 (32.6%)	711 (31.8%)
VTA	801 (22.3%)	631 (28.0%)	**1020 (33.1%)**	812 (36.3%)
Conditional	28 (0.9%)	27 (1.2%)	**79 (2.6%)**	68 (3.0%)
Imperative	28 (0.9%)	35 (1.1%)	**87 (2.8%)**	71 (3.2%)
kî- past	**714 (22.5%)**	573 (25.4%)	425 (13.8%)	307 (13.7%)
Quotative	**479 (15.1%)**	43 (1.9%)	249 (8.1%)	63 (2.8%)
	Of TAs…		**Of TAs…**	
LocalTA	12 (1.5%)	10 (1.6%)	**109 (10.7%)**	72 (8.9%)
MixedTA	**462 (57.7%)**	350 (55.5%)	490 (48.0%)	389 (47.9%)

Verbs, continued

	âcimisowina		kakêskihkêmowina	
	Tokens	Types	Tokens	Types
2Sg	149 (4.7%)	140 (6.2%)	**317 (10.3%)**	237 (10.6%)
1Pl	**518 (16.6%)**	427 (19.0%)	78 (2.5%)	68 (3.0%)
12Pl	52 (1.6%)	47 (2.1%)	**211 (6.8%)**	170 (7.6%)
2Pl	46 (1.5%)	41 (1.8%)	**146 (4.7%)**	126 (5.6%)
Unspecified	173 (5.5%)	150 (6.7%)	**290 (9.4%)**	239 (10.7%)
	Of quotatives…		**Of quotatives…**	
1Sg	100 (20.9%)	15 (34.9%)	**71 (28.5%)**	20 (31.8%)
3Sg	**451 (94.2%)**	29 (67.4%)	165 (66.3%)	24 (38.1%)
3Pl	17 (3.6%)	8 (18.6%)	**33 (13.3%)**	14 (22.2%)
Unspecified	9 (1.9%)	4 (9.3%)	**25 (10.0%)**	14 (22.2%)
Independent	**422 (88.1%)**	25 (58.1%)	158 (63.5%)	18 (28.6%)

Nouns

	âcimisowina		kakêskihkêmowina	
	Tokens	Types	Tokens	Types
Nouns	1851 (16.8%)	382 (18.8%)	1566 (15.3%)	489 (15.1%)
	Of nouns…		**Of nouns…**	
Singular	1124 (60.7%)	289 (42.4%)	**1159 (74.0%)**	304 (62.2%)
Plural	**525 (28.4%)**	211 (31.0%)	235 (15.0%)	97 (19.8%)
Obviative	**306 (16.5%)**	118 (17.3%)	180 (11.5%)	73 (14.9%)
Locative	**159 (8.6%)**	85 (12.5%)	39 (2.5%)	28 (5.7%)
Possessed	406 (21.9%)	164 (24.1%)	361 (23.1%)	152 (31.1%)
Dependent	359 (19.4%)	127 (18.7%)	253 (16.2%)	93 (19.0%)
	Of dependent…			
NID	**57 (15.9%)**	42 (33.1%)	19 (7.5%)	16 (17.2%)
NAD	303 (84.4%)	86 (67.7%)	**234 (92.5%)**	77 (82.8%)

Pronouns

	âcimisowina		**kakêskihkêmowina**	
	Tokens	Types	Tokens	Types
Pronouns	950 (8.6%)	83 (2.3%)	1430 (14.0%)	73 (2.3%)
	Of pronouns…		**Of pronouns…**	
Demonstr.	655 (69.0%)	37 (44.6%)	1159 (78.5%)	33 (45.2%)
Personal	175 (18.4%)	25 (30.1%)	146 (10.2%)	22 (30.1%)
	Of personal…		**Of personal…**	
First	81 (46.3%)	8 (32.0%)	73 (50.0%)	9 (40.9%)
Second	14 (8.0%)	5 (20.0%)	22 (15.1%)	77 (22.7%)
Third	69 (39.4%)	7 (28.0%)	47 (32.2%)	9 (40.9%)

Syntactic relations

	âcimisowina		**kakêskihkêmowina**	
ACT/GOAL	1237 (49.7% of nominals)		1519 (57.3% of nominals)	
	Of total…		**Of total…**	
Nouns		848 (68.6%)		878 (57.8%)
Pronouns		360 (29.1%)		635 (41.8%)
Dem + N	313 (16.9% of N,	47.8% of Dem)	345 (22.0% of N,	30.7% of Dem)

Linking features and functions

- *âcimisowina*
  - More first person exclusive
    - Stories about the speaker, likely not including the listener(s)
  - More quotatives
    - Retelling anecdotes about what other people said and did
- *kakêskihkêmowina*
  - More first person inclusive, more second person
    - Talking to and about the whole community
  - More imperatives
    - Directing the behaviour of the community
- (But there are other things that could also be influencing some of the features we looked at, such as age, gender, etc.)

UNIVERSITY OF ALBERTA

âcimowina & âtayôhkêwina

in the Bloomfield texts

UNIVERSITY OF ALBERTA

These groups were basically the same…

There were a few noticeable differences, but nothing like the texts we just looked at

Linking features and functions

- *âcimowina*
  - ???
- *âtayôhkêwina*
  - ???
- We could maybe learn something about the types of stories that fall into each of these categories, but only with much closer observation

UNIVERSITY OF ALBERTA

Discussion

Are the Bloomfield texts that much more homogeneous than the Ahenakew-Wolfart corpus?

- Maybe…

Or!

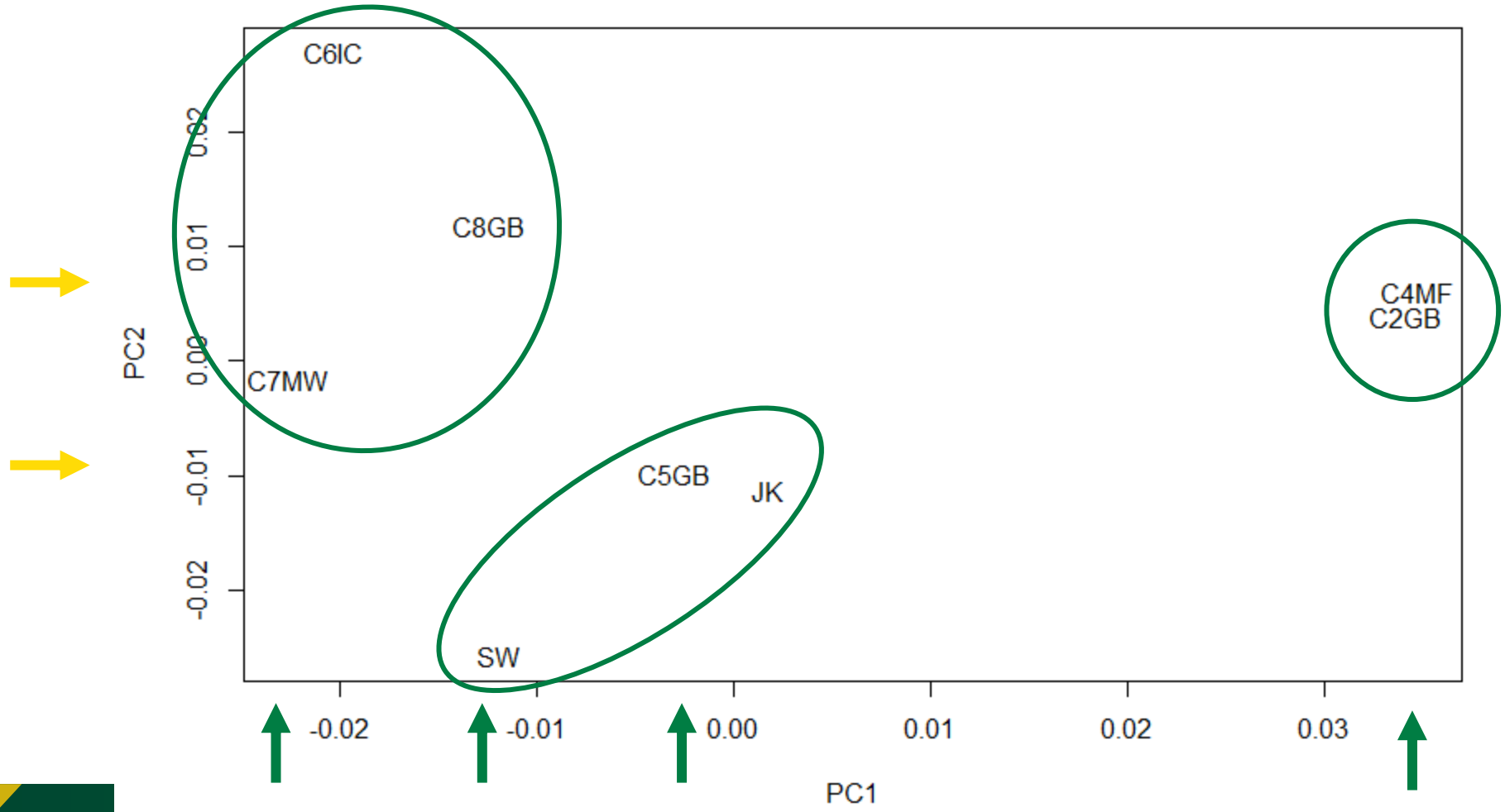
- We're looking at a broader distinction when we're looking at the Bloomfield texts
- Perhaps the overall differences between *âcimowina* and *âtayôhkêwina* are less apparent than the differences between various types of *âcimowina*
  - Last year, we looked at word order patterns in the two corpora, and the overall numbers didn't differ that much, but that changed when we looked closer at e.g. just VTAs (Schmirler & Arppe, 2019)

A top-down approach: the next steps

- Principal Component Analysis (PCA)
  - Lettings the texts group in terms of their features, rather than grouping the texts and then exploring their features
- Advantages: reduces complexity and collinearity
- Disadvantages: trickier to untangle what the outcome means from a human perspective

- We'll take a quick look at one example…

PCA


C6IC
C8GB
C7MW
C5GB
JK
SW
C4MF
C2GB
PC2
0.02
0.01
0.00
-0.01
-0.02
PC1
-0.02
-0.01
0.00
0.01
0.02
0.03

PCA

- PC1, top and bottom 10 (the features that best explain similarities/differences between the texts)
  - More positive: Quot, Ind, AI, Prs, 3Sg, Third, V, Prop, Imp, 3PlO
  - More negative: Ipc, A, Cnj, N, Prt, Sg, First, Pron, Pl, I
- We didn't look at particles earlier, so now we can:

	âcimisowina		**kakêskihkêmowina**	
	Tokens	Types	Tokens	Types
Particles	4739 (42.9%)	381 (10.5%)	3837 (37.5%)	427 (13.2%)
	Of particles…		**Of particles…**	
Negative	208 (4.4%)	12 (3.2%)	211 (5.5%)	10 (2.3%)
Locative	502 (10.6%)	103 (11.0%)	467 (12.2%)	31 (7.3%)
Temporal	354 (7.5%)	44 (11.6%)	252 (6.6%)	45 (10.5%)
Quantifiers	328 (6.9%)	30 (7.6%)	280 (7.3%)	34 (8.0%)

What did we learn?

- Using the bottom-up approach, we can find some differences between *âcimisowina* and *kakêskihkêmowina*
  - And link them to the situational context
- But no such luck for the broader categories we used for the Bloomfield texts
- Start with finer divisions of texts and then work to larger ones…
  - Even within the *âcimisowina* and *kakêskihkêmowina*, texts are far from uniform
    - A top-down approach shows how we might further refine the texts: cyclical method of register analysis!
  - Looking at *âcimowina* and *âtayôhkêwina*, we found basically nothing, and any links to situational context would be grasping at straws

- We have barely scratched the surface!

UNIVERSITY OF ALBERTA

ay-hay! Thank you!

Questions?

schmirle@ualberta.ca

arppe@ualberta.ca

References

Arppe, A., Schmirler, K., Harrigan, A. G., & Wolvengrey, A. (2020). A morphosyntactically tagged corpus for Plains Cree. Papers of the 49th Algonquian Conference.

Bear, G., Fraser, M., Wells, M., Lafond, A., & Longneck, R. (1998). kôhkominawak otâcimowiniwâwa // Our Grandmothers' Lives: As Told in Their Own Words (F. Ahenakew & H. C. Wolfart, Eds.; Bilingual edition). University of Regina Press.

Biber, D. (1991). Variation across Speech and Writing. Cambridge University Press.

Biber, D., & Conrad, S. (2019). Register, Genre, and Style (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108686136

Biber, D., Douglas, B., Biber, P. D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press.

Biber, D., Fitzmaurice, S. M., & Reppen, R. (2002). Using Corpora to Explore Linguistic Variation (Issue v. 9). John Benjamins Publishing Co. https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=253317&site=ehost-live&scope=site

Bloomfield, L. (1930). Sacred Stories of the Sweet Grass Cree. AMS Press.

Bloomfield, L. (1934). Plains Cree Texts. AMS Press.

Harrigan, A. G., Schmirler, K., Arppe, A., Antonsen, L., Trosterud, T., & Wolvengrey, A. (2017). Learning from the computational modelling of Plains Cree verbs. Morphology, 27(4), 565–598. https://doi.org/10.1007/s11525-017-9315-x

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

Schmirler, K., Arppe, A., Trosterud, T., & Antonsen, L. (2018). Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments. In N. C. (Conference chair), K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA).

Schmirler, K., & Arppe, A. (2019). Plains Cree actors and goals: Across time periods and genres. Paper presented at the 51st Algonquian Conference at McGill University in Montreal, Quebec, October 25-27, 2019.

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

Verbs

	âcimowina		âtayôhkêwina	
	Tokens	Types	Tokens	Types
Verbs	10298 (38.5%)	5483 (77.89%)	17968 (39.3%)	8449 (79.7%)
	Of verbs…		**Of verbs…**	
3Sg	5385 (52.3%)	2402 (43.8%)	10813 (60.2%)	4022 (47.6%)
1Pl	118 (1.2%)	103 (1.9%)	155 (0.9%)	128 (1.5%)
X	559 (5.4%)	400 (7.3%)	596 (3.3%)	425 (5.0%)
Quotative	954 (9.3%)	91 (1.7%)	2485 (13.8%)	149 (1.8%)
	Of quotatives…		**Of quotatives…**	
3Pl	69 (7.2%)	12 (13.2%)	81 (3.3%)	17 (11.4%)
Imperative	2 (0.2%)	1 (1.1%)	27 (1.1%)	8 (5.4%)

Nominals

	âcimowina		âtayôhkêwina	
	Tokens	Types	Tokens	Types
Nouns	5156 (19.3%)	1034 (14.7%)	8776 (19.2%)	1444 (13.6%)
	Of nouns…		**Of nouns…**	
Possessed	1679 (32.6%)	413 (39.9%)	3473 (39.6%)	663 (45.9%)
Dependent	1480 (28.7%)	297 (28.7%)	2976 (33.9%)	416 (28.8%)
	Of pronouns…		**Of pronouns…**	
Pers. pron.	227 (10.9%)	12 (25.5%)	340 (9.9%)	11 (22.9%)
	Of pers. pron.…		**Of pers. pron.…**	
Second	46 (20.3%)	4 (33.3%)	114 (33.5%)	4 (36.4%)
Third	100 (44.1%)	4 (33.3%)	101 (29.7%)	8 (36.4%)