

Contents of this talk

- 1. Background and Impetus
- 2. Methodology
 - a. Vector Space
 - b. Clustering
- 3. Results
- 4. Future Directions





Research Question



Can verbs be automatically semantically grouped in Plains Cree?



Background & Impetus

- -I'm just trying to finish my dissertation, okay?
 - Investigating morphological, syntactic, and semantic influences on alternations in Plains Cree order
 - Using a morphosyntactically inflected corpus (Arppe et al. in press)
 - But what about semantics?



Background & Impetus

-Plains Cree has multiple, high quality, dictionaries

-But it lacks an easily accessible semantic classification scheme (cf. Junker 2013)

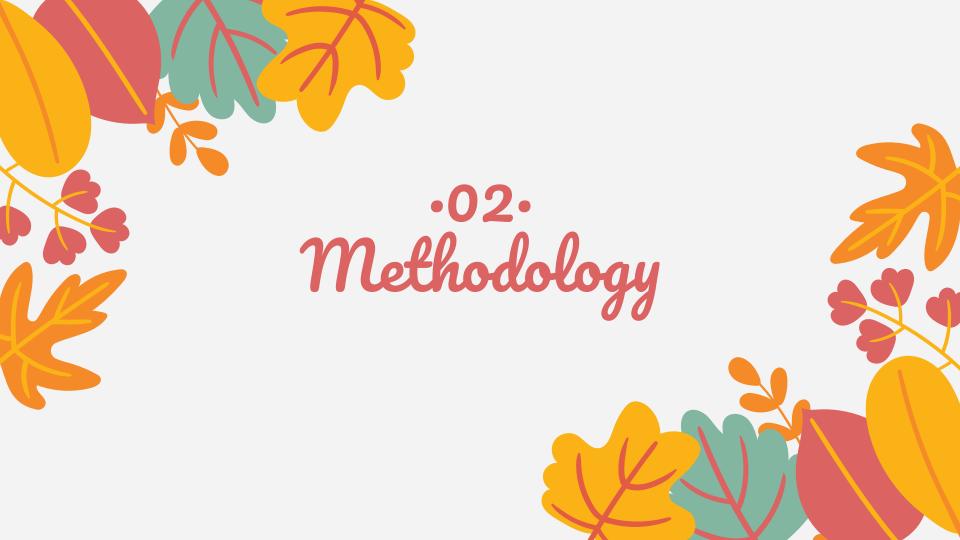
-Hand classification is time consuming and expensive.

- -Probably don't want to impose biases
- -How about those computers?

Background & Impetus

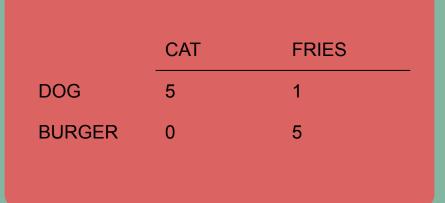
-Similar tasks performed by Divjak and Gries (2006) and Arppe (2008)

- Divjak and Gries (2006) cluster Russian TRY verbs based on morphosyntactic and semantic features to study structures of synonymy
- Arppe (2008) groups Finnish THINK verbs based on the other (similar) words used in their monolingual definition to investigate near synonymy



Vector Semantics

- -Based in the Distributional Hypothesis (Harris 1954; Firth 1957)
- -Uses co-occurrence matrices
- -Treats matrix columns as positional vectors in an n-dimensional space
- -Similar space = similar meaning



Vector Semantics (sparsity)

- -Zipfian distributions result in the majority of words occurring very few times
- -Resultantly, most word will not co-occur within a reasonable window of each other
- -Consequently, the majority of cells in a co-occurrence matrix will be 0, and vector spaces will be mostly empty
- -This can be mitigated by dimensionality reduction (e.g. Principal Components Analysis → neural nets → word2vec)

Clustering

- -Grouping together verbs that occupy similar spaces
- -Many methods to cluster, but we chose the basic Hierarchical Agglomerative Clustering (HAC) and the Ward method
- -Associated each lemma with its definition is a lexical database (Wolvengrey n.d.)
- -Cluster evaluation can take place!



Clustering

1. Using word-window and features

	Word 1	 Feat. 1	
Verb 1	0	2	
Verb 2	3	0	

2. Using just word-window

	Word 1	Word 2	Word 3	
Verb 1	9	1	8	
Verb 2	3	1	56	

Clustering

3. Using derivational morphemes

	Morph. 1		Morph. 2	
Verb 1		0	1	
Verb 2		1	0	



 Based on breakdowns from Wolvengrey's Database

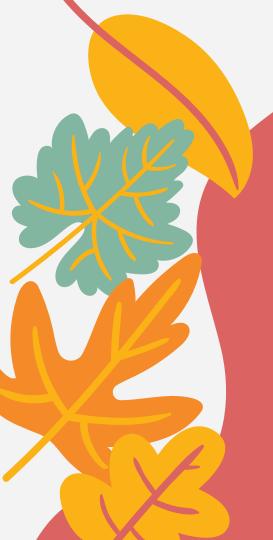




- -Some useful clusters
- -Not good enough to on its own
- -Decent scaffold







Word + Feature (VTA)

ayâwêw 15 s/he has s.o.

Itêw 14 s/he says thus to s.o.

Miswêw 13 s/he wounds s.o., s/he wounds s.o. with a missile

Isiyihkâtêw 12 s/he calls s.o. thus

Miyêw 11 s/he gives (it/him) to s.o.

Wîhtamawêw 11 s/he tells s.o. about (it/him)

Kanawêyimêw 7 s/he looks after s.o.; s/he takes care of s.o.

Osîhtamawêw 7 s/he makes (it/him) for s.o.; s/he arranges (it/him) for s.o.

osîhêw 7 s/he makes s.o. (e.g. bread); s/he arranges s.o.

otinêw 7 s/he takes s.o.; s/he takes s.o. in (e.g. orphan);

wîcihêw 7 s/he helps s.o.

wîcêwêw 7 s/he gets along with s.o.; s/he lives with s.o.

🤗 nipahêw 7 s/he kills s.o. 🤗



Word + Feature (VTA)

itâcimostawêw 5 s/he narrates thus to s.o.

itêyimêw 5 s/he thinks thus of or about s.o.

kakwêcimêw 5 s/he asks s.o.

kihci-wîkimêw 5 s/he marries s.o. formally

kiskêyimêw 5 s/he knows s.o.

kitotêw 5 s/he addresses s.o.

kitâpamêw 5 s/he looks at s.o.; s/he looks at s.o. with respect

kiyokawêw 5 s/he visits s.o.

miskawêw 5 s/he finds s.o.

nakatamawêw 5 s/he leaves (it/him) behind for s.o.;

nakayâskawêw 5 s/he is comfortable with s.o.

sâkihêw 5 s/he loves s.o.

nitohtawêw 5 s/he listens to s.o.

nitotamawêw 5 s/he asks s.o. for (it/him)



Word + Feature (VTA)

cimahêw cîhkêyimêw

cîsiskawêw

isinawêw

ispisîhêw

itakimêw

itinêw

itisahamawêw

itisinamawêw

itwahwêw

itôtawêw

kapêsimostawêw

katawatêyimêw

4 s/he places s.o. (e.g. tree) upright

4 s/he likes s.o.

4 s/he flirts with s.o. by slight body contact

4 s/he sees s.o. thus; s/he has a vision of s.o.

4 s/he raises s.o. thus

4 s/he counts s.o. thus; s/he value s.o.

4 s/he moves s.o. thither or thus by hand

4 s/he sends (it/him) to s.o. thus or there

4 s/he hands (it/him) over to s.o.

4 s/he points his/her finger at s.o.

4 s/he does thus to s.o.

4 s/he camps overnight near s.o.

4 s/he thinks s.o. beautiful



Word + Feature (VII)

pêyakwan 8 same the same just the same; similar

isiyihkâtêw 7 s/he calls s.o. thus s/he gives s.o. such a name;

miywâsin 6 good thing

ispayiw 6 it comes by it goes by; it runs thus (in a cycle)

ayâw 5 s/he is s/he is there

pipon 4 it is winter; is one year

nîpin 4 it is summer

kîsikâw 4 day daylight; day sky

astêw 3 it is there it sits there

ihtakon 3 it exists it is there

âpatan 3 it is used it is useful

âyiman 3 it is difficult



Word + Feature (VAI)

ispayiw 12 it takes place thus it occurs thus isiyihkâsow 12 s/he is named thus s/he has such a name astâw 12 s/he puts s.t. there s/he places s.t. there itwêw 11 s/he says so ayâw 9 s/he is s/he is there

wîkiw 8 s/he lives there

takohtêw 8 s/he arrives s/he arrives walking

sipwêhtêw 8 s/he leaves s/he goes off s/he departs

pimâtisiw 8 s/he lives s/he is alive; s/he survives

nîmihitow 8 s/he dances

kiskisiw 8 s/he remembers

itohtêw 8 s/he goes there

atoskêw 8 s/he works



Word + Feature (UTI)

```
kiskêyihtam 10 s/he knows s.t.
itêyihtam
              9 s/he thinks thus of or about s.t. s/he regards s.t. Thus
              8 s/he uses s.t. s/he makes use of s.t.
âpacihtâw
wîhtam
              8 s/he tells about s.t.
wâpahtam
              8 s/he sees s.t. s/he witnesses s.t.; s/he has a vision
otinam
              8 s/he takes s.t.; s/he chooses s.t.
ohtisiw
              8 s/he receives s.t. from there
nitawêyihtam 8 s/he wants s.t. s/he desires s.t.
kîsihtâw
              8 s/he completes s.t. s/he finishes s.t.
kisîpêkinam
             8 s/he washes s.t. (e.g. clothing)
              8 s/he says thus to or about s.t.; s/he calls s.t. thus
itam
             8 s/he calls s.t. thus
isiyihkâtam
              8 s/he puts s.t. there s/he places s.t. There
astâw
              7 s/he tells about s.t. s/he gives an account of s.t.
âtotam
              6 s/he does s.t. s/he does s.t. so s/he acts thus
tôtam
osîhtâw
              6 s/he makes s.t. s/he prepares s.t.
```

Just Words (VAI)

isiyihkâsow 12 s/he is named thus s/he has such a name

Itwêw	11	s/he says so
nîmihitow	10	s/he dances
kiskisiw	10	s/he remembers
atoskêw	10	s/he works
âstawêhikê	w 9	s/he extinguishes the fire
wanisin	9	s/he is lost
tâpwêw	9	s/he speaks true s/he speaks the truth
takosin	9	s/he arrives
takopayiw	9	it arrives (e.g. a vehicle)
pîkiskwêw	9	s/he speaks
pimohtêw	9	s/he walks



Just Words (UTI)

isiyihkâtêw	15	s/he calls s.o. thus
ayâwêw	14	s/he has s.o.
itêw	13	s/he says thus to s.o.
âpacihêw	12	s/he uses s.o.
wîcihêw	12	s/he helps s.o.
tipahamawêw	12	s/he pays (it/him) to s.o.
pakitinêw	12	s/he lets s.o. Go
osîhêw	12	s/he makes s.o. (e.g. bread)
osîhtamawêw	12	s/he makes (it/him) for s.o.
oswêw	12	s/he brings s.o.
nôcihêw	12	s/he hunts s.o.
nitomêw	12	s/he invites s.o.
mowêw	12	s/he eats s.o.
miswêw	12	s/he wounds s.o.



Derivational Morphemes (VII)

âhkwatin 8 it is freezing

takwâkin 8 it is autumn, it is fall

nîpin 8 it is summer

nistinwa 8 they are three in number

namatakon 8 it is absent

mispon 8 it snows, it is snowing, there is falling snow, there is a snow-fall

mihcêtinwa 8 they are numerous, they are many, they are plentiful

ihtakon 8 it exists, it is there

ihkin 8 it happens, it occurs, it takes plac<mark>e</mark>

akohtin 8 it soaks, it is in water; it floats on liquid (e.g. grease on soup)



yôhtênikâtêw	6	it is	open	it	is op	ened
wêpinikâtêw	6	it is	throw	'n	away	/

takonikâtêw 6 it is added in by hand

tahtinikâtêw 6 it is taken off it is loosened it is untied

sisoskiwakinikâtêw 6 it is plastered it is mudded

pîkonikâtêw 6 it is broken it is torn it is broken down

pitikonikâtêw 6 it is shaped to be round

otinikâtêw 6 it is taken it is obtained; it is chosen; it is purchased

kisîpêkinikâtêw 6 it is washed

asênikâtêw 6 it is refused it is rejected; it is turned down it is turned back



wîci-ohpikîmêw s/he grows up together with s.o. wîci-kiskinohamâkosîmêw 5 s/he is in school with s.o. wîci-ayâmêw s/he lives with s.o. s/he dwells with s.o. wîc-îspîhcisîmêw s/he is of the same age as s.o. wîc-âyamihâmêw s/he prays with s.o. wêmistikôsîmototawêw s/he speaks French to s.o. pôsihêw s/he gives s.o. a ride paci-tôtawêw s/he wrongs s.o. mâyi-tôtawêw s/he does a bad thing to s.o. miyo-âpacihêw s/he uses s.o. well s/he makes good use of s.o. miyo-wîcêwêw s/he gets along well with s.o.

Conclusion

- -Automatic classification sorta works
- -Word + Features more useful than word window alone
- -Can work as a scaffold on which to build classes
- -Most frequent items on their own
- -More data would likely help

Future Research

- Add more data to the corpus
- Levenshtein distance?
- Word2Vec
- Different clustering and dimensionality reduction methods
- Have Native Speaker(s) look at clusters (without definitions)

Thanks!

Does anyone have any questions?

galvin@ualberta.ca

a.ttic.us

arppe@ualberta.ca

https://sites.ualberta.ca/~arppe/english.html

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

References

Arppe, A. (2008). Univariate, Bivariate and Multivariate Methods in Corpus-Based Lexicography - a Study of Synonymy. (PhD Dissertation). University of Helsinki, Helsinki, Finand.

Arppe, A., Schmirler, K., Harrigan, A. G., Wolvengrey, A. (in press). Morpholosyntactically Tagged Corpus For Plains Cree in Macaulay M. & Noodin M. (Eds.), Papers of the Forty-Ninth Algonquian Conference. EAST LANSING: Michigan State University Press.

Divjak, Dagmar & Gries, Stefan. (2006). Ways of trying in Russian: clustering behavioral profiles. Corpus Linguistics and Linguistic Theory - CORPUS LINGUIST LINGUIST THEO. 2. 23-60. 10.1515/CLLT.2006.002.

Firth, J. (1957). A Synopsis of Linguistic Theory. Oxford: Blackwell.

Harris, Z. S. (1954). Distributional structure. Word, 10 (2-3), 146–162.

Junker, M. O. (2013). East Cree (Northern) Thematic Dictionary. Cree School Board & Eastcree.org.