Department of Political Science
University of Toronto Scarborough
m.cowan@mail.utoronto.ca

Stagnant Promises

Analyzing Treaty Dynamics in Canadian-Indigenous Relations

Cowan, M., Cowie, C., Greenaway C., & Schnurbusch, P. (n.d.)

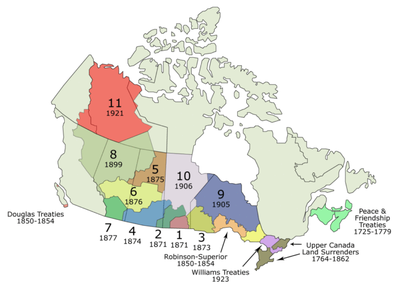
UNIVERSITY OF TORONTO SCARBOROUGH

Introduction

The historical relationship between Indigenous nations and the Canadian state has been shaped by forms of Indigenous-European diplomacy. Treaties are living documents that continue to influence legal, political, and social landscapes in Canada—with implications for land rights, resource management, and the sovereignty of Indigenous peoples that affecting contemporary policy-making and reconciliation efforts.

Spirling (2012) employed natural language processing (NLP) and string kernel principal components analysis (KPCA) to explore how lexical changes in U.S.–Indian treaty language reflected the country's relative bargaining power, contributing to worsening terms for Indigenous Nations over the 19th century. We apply KPCA to Crown-Indigenous treaties, contextualizing the interplay between colonial objectives and Indigenous self-determination within a two-century historical framework.

Treaty Sample (*n* = 59)

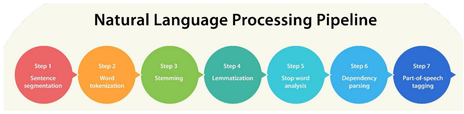


Treaties	Sample
Treaties of Peace & Neutrality (1701)	1
Peace & Friendship (1725–1779)	4
Upper Canada Land Surrenders (1763–1862)	25
Royal Proclamation (1763)	1
Robinson Treaties (1850)	2
Douglas Treaties (1850–1854)	13
The Numbered Treaties (1871–1921)	11
Williams Treaties (1923)	2

NLP: from Words to Numbers

banana [0.5 0.12 ... 0.23]

1. **Bag-of-Words:** Count only the frequency of the words (ignores word order).
   - "the" = 2
   - "quick" = 1
   - "brown" = 1
   - "fox" = 1
   - "jumps" = 1
   - "over" = 1
   - "lazy" = 1
   - "dog" = 1
2. **N-grams:** Captures small ***n*** sequential words or characters (e.g., *n* = 2).
   - "The quick"
   - "quick brown"
   - "brown fox"
   - "fox jumps"
   - "jumps over"
   - "over the"
   - "the lazy"
   - "lazy dog"
3. **String Kernels:** Calculates similarity using all sequences within a specified length (e.g., length = 5).
   - "The q"
   - "he qu"
   - "e qui"
   - "quic "
   - "quick"
   - "uick "
   - "ick b"
   - "ck br"



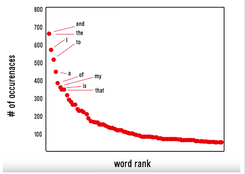
Term Document Matrix with TF-IDF Weighting

Tokenization is the process of breaking text into individual tokens (words, characters, or sub-words) when we have:

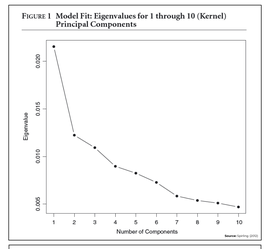
- Standardized (e.g., lowercase, remove accents);
- Removed noise (e.g., *stop words*, sparse terms, !@#$%^&*);
- Reduced dimensionality (e.g., stemming, lemmatization);
- **Goal:** Extract only meaningful features.

TF-IDF weighting adjusts term frequencies by penalizing common terms that appear across many documents and highlighting terms that are significant within specific documents.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log\left(\frac{N}{\text{DF}(t)}\right)$$

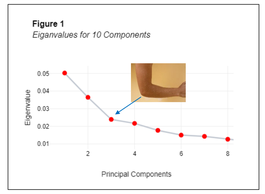


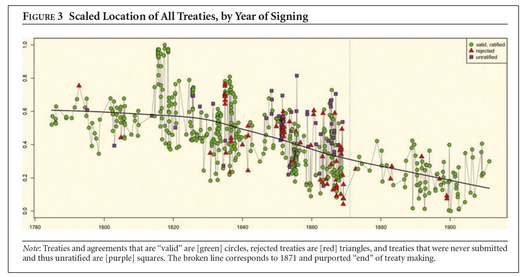
String Kernel Principal Components Analysis



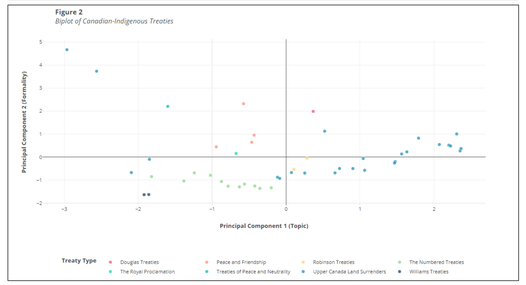
Eigenvalues: indicate the amount of variance in the data that is explained by its corresponding principal component.

- Higher eigenvalues mean that the principal component captures more variance from the data.





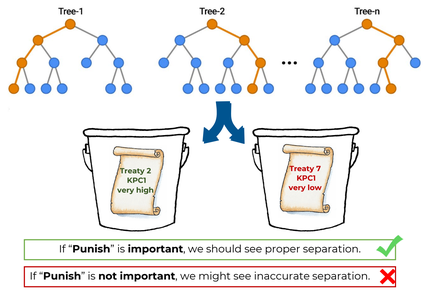
Note: Treaties and agreements that are "valid" are [green] circles, rejected treaties are [red] triangles, and treaties that were never submitted and thus unratified are [purple] squares. The broken line corresponds to 1871 and purported "end" of treaty making.



Random Forest (Word Importance) & Correlations

Random Forest Algorithm: think of the entire sample of the treaties as the "forest".

- The algorithm plants **word-stem "trees"** at random points in the corpus.
- Each tree makes a predictive "bucketing" decision extending throughout the entire corpus before the algorithm aggregates its calculations.



Word stems with a positive correlation appear more frequently/are more prominent in treaties that have high scores on the principal component.

Table 1
KPC1 Correlations: Land Transactions vs. Governance Agreements

Term	Frequency	Correlation
unto	27	0.7871176
successor	34	0.7844702
receipt	14	0.6657561
grant	29	0.6555149
sell	20	0.6299012
parcel	19	0.6184527
tract	36	0.5273126
dominion	14	-0.3754310
surrend	28	-0.3898219
punish	13	-0.4135275
govern	19	-0.4335618
protect	11	-0.4909522
resid	12	-0.5009835
indian	45	-0.6514852

Note:
Terms in KPC1 reflect transactional language and governance, with high correlations for property rights and low correlations for governance themes.

Word stems with a negative correlation appear more frequently/are more prominent in treaties that have low scores on the principal component.

Table 2
KPC2 Correlations: Formality and Legal Language

Term	Frequency	Correlation
great	37	0.6968075
father	14	0.6912476
children	18	0.6818651
protect	11	0.6493388
white	20	0.5985758
cultiv	15	0.5786606
thought	10	0.2771740
stipul	13	-0.4032312
offend	13	-0.4196652
pursuant	12	-0.4607742
negoti	13	-0.4710283
infring	12	-0.4805473
commission	13	-0.5418299

Note:
Terms in KPC2 highlight the level of formality and hierarchical legal language, contrasting formal legal discourse with more relational language.

US-Canada Comparison

Table 3
Comparison of Stem Correlation Scores for U.S. and Canadian Treaties

	Correlation		
Word	U.S. Treaties	Canadian Treaties (KPC1)	Canadian Treaties (KPC2)
friendship	0.62	-0.51	0.37
peace	0.56	0.32	0.04
mutual	0.49	0.03	0.14
boundary	-0.01	-0.04	-0.10
relinquish	-0.08	-0.27	0.15
dollar	-0.10	-0.32	-0.49
tract	-0.20	0.52	-0.27
land	-0.49	0.11	0.03
reservation	-0.49	-0.32	-0.50

Note:
Correlation scores for selected stems in U.S. treaties (Spirling, 2012) and Canadian treaties (KPC1 and KPC2). Spirling's analysis identified patterns of tonal shifts in U.S. treaties, where positive and negative correlations reflected changes in relational tone over time. In contrast, the Canadian treaty correlations are inconsistent across KPC1 and KPC2, showing no unified pattern that would indicate tonal shifts. This suggests a lack of relational or conflict-driven language dynamics in Canadian treaties, reflecting instead a stable, administrative language that does not respond to historical context in the same way.

- We contend that the structure of **Canadian/British approach to treaty-making was formulaic**—unlike the U.S. model, which exhibited greater variation in response to historical context.
- The British, and subsequently **Canadian, authorities employed a standardized legal language** that reinforced their colonial objectives—an approach that contrasts with the diverse negotiators, state-to-state variations, and greater linguistic diversity found in the U.S. context.
- The fact that **treaties cluster based on thematic rather than temporal-event criteria** supports the notion that Canadian treaties were not particularly responsive to historical events or shifts in institutional dynamics—instead, exemplifying a stable treaty-making framework entrenched in broader colonial governance.

Validation

1. Qualitative

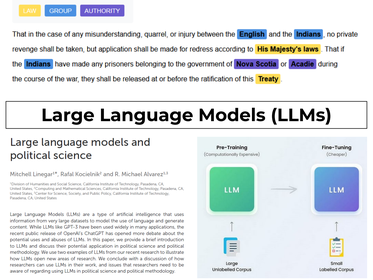
Search function

Subject-matter expertise

2. Quantitative

Root words

Annotation



Large Language Models (LLMs)

References

1. Linegar, M., Kocielnik, R., & Alvarez, R. M. (2023). Large language models and political science. *Frontiers in Political Science, 5*. https://doi.org/10.3389/fpos.2023.1257092
2. Spirling, Arthur. 2012. "U.S. Treaty Making with American Indians: Institutional Change and Relative Power, 1784-1911." *American Journal of Political Science 56* (1): 84–97. https://doi.org/10.1111/j.1540-5907.2011.00558.x.

Acknowledgements

- Join the UTSC Methods Hub discord: https://discord.gg/N4edY2J5
- Scan the QR code to visit the SDAC GitHub.
- Look for CCR-Accredited Data Workshops in Fall 2025.