

DIGI405-24S1 Corpus Analysis Project – pkw21

Topic: National vs. Labour on Gang Discourse. Does Labour use “softer” language?

Kane Williams [pkw21@uclive.ac.nz]

13th May 2024

1. Introduction

This paper performs a (1) keyness analysis, (2) collocation analysis, and a (3) concordance on a subset of New Zealand parliamentary speech data. The goal is to determine if and how Labour and National use different language when discussing Gangs. It will be shown that there is weak evidence for a “Labour-Soft” thesis: that is, National tends to use harsher language, such as “violence” when referring to gangs, and Labour tends to use softer language, such as “community”.

Outline

1. Introduction.....	1
2. Background.....	2
3. Data and Methods	2
4. Results and Discussion	3
5. Conclusion	8
6. Appendix	9
7. References.....	9

2. Background

Christopher Luxon is the leader of New Zealand's *National* party, and the current prime minister of New Zealand. Before coming into power, he accused the previous government (led by *Labour*) of being too soft on gangs (RNZ, 2022). In response, Luxon as prime minister has said he wants to 'crack down' on gangs¹, and his party has recently enacted legislation with that aim (NZ Legislation, 2024).

Williams' et. al. (2024) generated a corpus of politician's speeches using New Zealand parliamentary transcripts in the form of "Hansard Reports" (NZ Parliament, 2024). The hypothesis Williams' et al. built the corpus to test (Williams, 2024, p.1) was whether "Labour is soft on gangs" can be revealed in the manners of speech different party members use when discussing gangs.² Call this the *Labour-Soft* thesis.

3. Data and Methods

The data consists of web-scraped New Zealand parliament transcripts, filtered to only the dialogues containing at least one of "gang*", "mongrel mob", and "black power". The transcripts are then separated by speaker, and finally sorted by party.

There are 585 files sorted into the *National* corpus, with a total word count of 384,329. There are 348 files in the *Labour* corpus, with a total word count of 247,191. The data is from 2003 to 2024. While Luxon made his claim about the 2017-2023 Labour government particularly, this larger set of data will nevertheless provide more information to work with. Further information on the dataset can be found in (Williams, 2024).

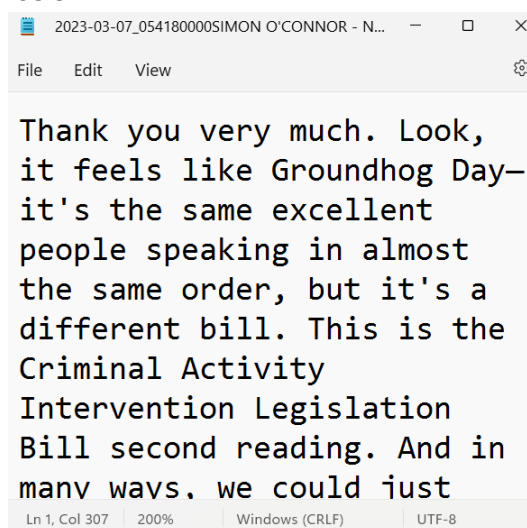


Figure 1 – A typical file. This is from Simon O'Connor, and has been sorted into the "National" corpus.

All the following methods were applied using AntConc 3.5.9 running on Windows. For a visual overview of the workflow, please consult *Figure A* in the Appendix.

Keyness

To identify different patterns of speech, a natural first step is to apply a **Keyness** analysis, using the National corpus as the target corpus, and the Labour corpus as the reference corpus. This will pick out words that are distinct in one corpus relative to the other.

If the *Labour-Soft thesis* holds, we might expect the following types of words, emphasizing the humanity of gangs and working together with them, to be more prominent in the Labour corpus compared to the National corpus:

¹ For more on Gangs and New Zealand's political response, see (Gilbert, 2013).

² Luxon is almost certainly not making a claim about how politicians talk, but rather about government policy and behaviour toward gangs. Nevertheless, his claim makes an interesting starting point for a discourse analysis project.

Rehabilitate, Collaborate, Together, Support

Likewise, we might expect the following types of words, which emphasize the criminality of gangs and point to punishments, to be more prominent in the National corpus:

Punish, Law, Criminal, Violent

Collocations

Apart from picking out words used more frequently throughout the entire corpus, another natural place to begin is with an analysis of words that collocate with the word “gang” in both the National and the Labour corpora separately. (Caldas-Coulthard, 2010) provide an example of how collocations can reveal “societal attitudes”, so we might expect a similar set of words to appear as those mentioned in Keyness above.

Concordance

After finding the words that appear relevant to the *Labour-Soft* thesis using keyness and collocation analyses, a concordance analysis can be applied to observe if these keywords really are relevant to the thesis.

This is important in the context of parliamentary discourse, because “the nature of parliamentary discourse is not monologic but dialogic” (Bayley, 2004, p.25). This means that speakers of one party may not be speaking their own views, but rather speaking about what members of other parties are saying.

For example, if Labour uses “Support” more than National, it may be because they are making claims of National such as “National’s policy supports gangs, while our policy does not.” This says little about Labour’s own views towards gangs. Therefore, it is important to keep in mind the specific context of the words being used when performing a keyness or a collocation analysis.

Occasionally, the following two methods will also be used to complement the concordance:

- Counting the occurrences of an n-gram, and
- Viewing the specific files themselves.

4. Results and Discussion

Keyness

First, the Keyness Analysis was performed, with National as the target corpus and Labour as the reference corpus. For all keyness analyses, a Dice coefficient was used as the effect measure, at the $p < 0.05$ level with the Bonferroni correction applied. Recall that the Dice coefficient measures the overlap of words – words appearing in both corpora - with a 1 suggesting a total overlap, and 0 saying there is no overlap.³

³ See (Rychly 2008) for more on using the Dice Coefficient and alternatives.

No stop word or lemma list was used initially. For stoplists, this was because there might be information contained in words typically appearing in stoplists. For example, a party might use “us” or “they” more often if they are focused on characterising an in-group and an out-group, which has the potential to be relevant. Likewise, information might be lost if words are combined via a lemma list.

A Positive keyness scores mean that National uses these words more frequently (with Labour as the reference). Negative keyness scores mean the opposite.

Table 1: Relevant words in the top 10 - Ordered by Keyness – Stop Words Included									
National (Rank: 1 [High] - 65 [Low])					Labour (Rank: 66 [High] - 120 [Low])				
Rank	Freq	Keyness	Effect	Keyword	Rank	Freq	Keyness	Effect	Keyword
2	3870	+ 128.78	0.0194	they	66	81	- 249.51	0.0004	māori
9	199	+ 63.3	0.001	violent	68	175	- 105.18	0.0009	organised

Table 1 above shows the noticeable words from the top 10 keyness results for both National and Labour. The inclusion of “Maori” will be discussed shortly. This list is limited, so the search will be expanded by adding a stopword list⁴ (to filter out likely the likely irrelevant words dominating top 10 such as “s”, “re”), and considering words beyond the top 10 of either list.

Table 2: Relevant words - Ordered by Keyness – Stop Word Excluded									
National (Rank: 1 [High] - 53 [Low])					Labour (Rank: 54 [High] - 107 [Low])				
23	53	+ 30.98	0.0006	shootings	Rank	Freq	Keyness	Effect	Keyword
27	56	+ 30.2	0.0006	shooting	105	213	- 21.64	0.0023	communities
29	66	+ 29.64	0.0007	rehabilitation					
30	154	+ 28.92	0.0017	drug					

Table 2 above shows a few more words that are potentially relevant. Figure 2 to the right shows the top 10 words for Labour. Note that $\frac{5}{10}$ of the words are in the Māori language, nearly all with a very low dice coefficient.

Labour has historically been more appealing to Maori voters than Labour. Furthermore, more Labour Members of Parliament have been Maori compared to National, historically. It is possible that this difference could contribute to different attitudes towards gangs, who tend to be

Rank	Freq	Keyness	Effect	Keyword
54	81	- 249.62	0.0009	māori
55	48	- 167.47	0.0005	te
56	175	- 105.23	0.0019	organised
57	1	- 81.61	0	garrett
58	325	- 67.9	0.0035	act
59	17	- 61.23	0.0002	e
60	29	- 60.74	0.0003	borrows
61	7	- 58.61	0.0001	kia
62	3	- 58.27	0	quinn
63	225	- 57.83	0.0024	problem
64	6	- 57	0.0001	ki

Figure 2 - Words with a high Keyness score for Labour. Stopwords excluded. Note the high proportion of Māori words.

⁴ The NLTK-English stopwords list provided in the first week of this course was used. (Grimmer, 2013) recommends removing words that appear in more than 99% of documents, and less than 1%, in political texts (depending on context). Given more time, this would have been another strategy to try.

predominantly Maori. Much more analysis would need to be done to investigate these claims, however whether it is correct will certainly influence the truth of the Labour-soft thesis

Collocation

Next, a Collocation Analysis was performed on the word “gang”, on both the National and the Labour corpus separately.

A window span of 5L to 5R was used. A minimum collocation frequency of 1 was found be the most informative, as using a higher frequency causes the lists to start looking similar. A mutual information score was used as the statistic, and the lists were ordered by this statistic. Recall that a mutual information score measures how “surprising” it is to see certain words besides each other, potentially highlighting a strong connection.

The top 13 collocates of “gang” sorted by mutual information score are shown below in Table 3 below.

Table 3: Top 13 Collocates for “gang” – ordered by Mutual Information			
National		Labour	
Stat	Collocate	Stat	Collocate
9.09920	thuggish	9.08800	pads
9.09920	identifiable	9.08800	accumulation
9.09920	headhunter	8.67296	affiliations
9.09920	friendship	8.08800	willingness
9.09920	exacerbate	8.08800	whanau
9.09920	eventuate	8.08800	vanish
8.68416	rivalries	8.08800	undress
8.51423	raped	8.08800	trophy
8.51423	argues	8.08800	troopers
8.32159	forts	8.08800	treats
8.09920	yelled	8.08800	transforming
8.09920	wears	8.08800	thirds
8.09920	visibly	8.08800	tawhero

Noticeable words for National include “thuggish”, “friendship”, “rivalries”, “raped”, “argues”, and “yelled” (Note that “headhunter” refers to the name of a New Zealand gang.)

Noticeable words for Labour include: “whanau” (roughly “Family”, in the Maori language), and possibly “troopers”, “treats”, and “undress”.

From both the keyness and the collocation analysis, we have the following list of words (Table 4) that could be investigated with both a concordance and a deeper investigation into the specific texts themselves:

Table 4: Keywords - Words in the Corpus possibly relevant to the Labour-Soft thesis.	
National	Labour
<ul style="list-style-type: none"> - They - Violent - Shooting(s) - Rehabilitation - Drug - Thuggish - Friendship - Rivalries - Raped - Argues - Yelled 	<ul style="list-style-type: none"> - Maori - Organised - Communities - Whanau - Troopers - Treats - Undress

A cursory glance at these words paints a picture of National possibly associating gangs with barbarism and brutality – with words such as “thuggish”, “violent”, and “rape”. At the other end, Labour appears to associate gangs with an organised family-oriented structure – with words such as “Organised”, “Communities”, and “Whanau”.

Concordance

Due to time and space constraints, particular attention will be paid to only the two **Bold** words in Table 4.

Organised

Using a Kwic sort of 1R, 2R, and 3R – to pick up the nouns that “organised” modifies:

Table 5: Example of concordances of “organised”, along with counts.	
National	Labour
Concordance Hits 175 Hit KWIC 1 et papers, the new Organised and Financial Crime Agency is sup 2 ings and very well- organised and well-attended —possibly beca 3 s become far more organised around it, and we see treaties like 4 ds. I had a meeting organised by Destination Orewa and Hellen 5 / police focused on organised crime ? 6 combat gangs and organised crime ? 7 rms and drivers of organised crime . 8 ng leaders, and the organised crime administrators; those people 9 nouncement of the Organised Crime Agency in a half-baked stat 10 regarding the new Organised Crime Agency she has mentioned 11 take steps against organised crime and criminal gangs. I find it i 12 d been involved in organised crime and criminal gang activity. It 13 ad of incidences of organised crime and gang activity in New Ze	Concordance Hits 292 Hit KWIC 93 roach to gangs and organised crime, but just again emphasising 94 ment to respond to organised crime by ensuring the enforceme 95 sed, but the words “ organised crime ” certainly are used quite a l 96 at the transnational organised crime conference, where groups f 97 lthough there were organised crime connections. We now know 98 ces, participation in organised crime, criminal groups —so some 99 those gangs and by organised crime diminish, but I guess our vi 100 be able to deal with organised crime effectively. This is not an ea 101 gang members and organised crime—far greater than any achie 102 cifically focused on organised crime; far more than when that m 103 ho are dedicated to organised crime-fighting, and the total polic 104 uch as members of organised crime fraternities to own firearms 105 full range to tackle organised crime from the beginning to the e

The word “organised” in both corpora is overwhelmingly used as part of the noun phrase “organised crime”. Furthermore, there is a clear difference in how often this noun phrase is used in National corpus compared to the Labour corpus.

In the National corpus, “organised cri*” has 89 hits out of 585 documents. That is, it occurs in $\frac{89}{585} = 15\%$ of all documents.

In the Labour corpus, “organised cri*” has 117 hits out of 348 documents. That is, it occurs in $\frac{117}{348} = 34\%$ of all documents.

Looking at the Labour text files individually shows us how “organised crime” is used:

...particularly with reference to this bill, if a person is a member or has close affiliation with a gang or **organised crime** group. ... there's been a lot of discussion around the current Government's approach to gangs and **organised crime**, but just again emphasising... [2021-10-20_0527400000LIZ CRAIG.txt]

...huge weapon in the fight against **organised crime** and gangs... key tools to attack gangs and **organised crime**. Again I say that I mean no disrespect, but I think that this bill is a substantive measure ... and dealing with gang leaders and **organised crime**. Some of the most insidious gangs and elements of **organised crime** in our community... [2009-03-10_00001195CLAYTON COSGROVE.txt]

The phrase “organised crime” is often used to complement “gang”, or to refer to gangs. This is consistent throughout the files which contain “organised crime” in the Labour and the National corpus.

That 34% of Labour files contains “organised crimes” compared to National’s 15% is very noticeable. A hypothesis is that Labour is trying to position gangs as being non-thuggish, in contrast to National. Gangs would be more akin to the Mafia, who can be reasoned with, than Barbarians, who must be punished. This could be evidence for the Labour-soft thesis – if it is true it would give Labour reason to be “soft” on gangs.

As compelling as this view is, there is a limitation. Firstly, even if it is true that Labour uses the words “gang” and “organised” together this might have no bearing at all regarding how Labour members conceptualize gangs, any more than it is true that someone who often says “heavy rain” thinks about the weight of rain. It may just be an unconscious manner of speech, and a phrase (“gangs and organised crime”) commonly used within internal Labour discussions, but not internal National discussions, that has propagated to the parliament floor.

Nevertheless, this stark difference in the frequency of the phrase in the corpora suggests possible evidence that the Labour-soft thesis holds.

Whanau

There is only one file in either corpus that uses “whanau”, 2022-04-13_0593400000VANUSHI WALTERS.txt. This is a 2022 text sorted into the “Labour” category with Vanushi Walters as the speaker and it ranks with a very high “mutual information” score for “gang”. In it, she is quoting an article on the Salvation Army’s website⁵. What follows is the relevant part of the text file:

⁵ <https://www.salvationarmy.org.nz/article/challenging-prejudice>

...editors wrote in part about gangs, published on their website: "we must remember that New Zealand gangs have their origin in urban poverty and disconnection ... the greatest danger of prejudice is that it excludes people, pushing them to the margins. Once situated there, it becomes even easier for us to treat them as outcasts. But plenty of those who grew up in gangs don't want that for their children. As Pat of the Wellington Black Power says in this edition, 'Most of our **whanau** are born into [the gang], and my role is to improve our kids' way of thinking and steer them away from the gang scene. This is the change I'm trying to make.'"...

Here, Vanushi quotes the Salvation army, who are themselves quoting a Black Power member, so it may almost seem unreasonable to attribute this usage of "whanau" to the Labour party. Nevertheless, the fact that she is trying to highlight this article and this quote – which emphasizes the poverty and humanity of gangs - might nevertheless be construed as a very (since it comes from only one politician) weak form of evidence towards the Labour-soft thesis.

One possible limitation to all the above analyses is that the data varies over different governments and time periods. The manner of speaking may be different in 2003 than it 2023. Furthermore, gangs may be a bigger problem in some years more than others, which means that certain years – run under specific governments - will have a lot of data, and certain years will have little data. For more on this limiting aspect of parliamentary data, see (Ivanush, 2023). Given more time, it could be informative to perform each of the above analyses but segregated by government.

5. Conclusion

This paper has investigated the "Labour-Soft" thesis: that National and Labour members use different language when discussing gangs. Using Keyness between the two corpora, a collocation analysis on "gang", and a concordance and frequency count on the keywords previously found, it has been shown that there are multiple sources of very weak to weak evidence that the thesis holds.

WORD COUNT (excl. quotes, tables): < 2180.

6. Appendix

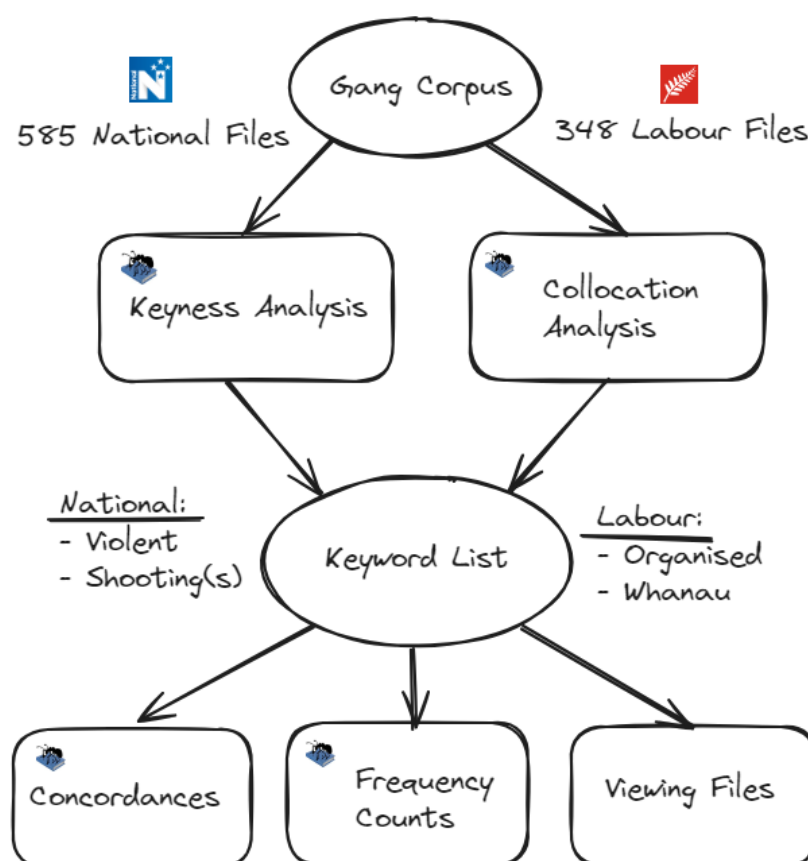


Figure A – Workflow and order of analyses.

7. References

Bayley, P. "The whys and wherefores of analysing parliamentary discourse." In P. Bayley (Ed.), *Cross-cultural perspectives on parliamentary discourse*. Amsterdam/Philadelphia: John Benjamins Publishing, (2004).

Caldas-Coulthard, C. R., & Moon, R. (2010). 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society*, 21(2), 99-133.
<https://doi.org/10.1177/0957926509353843>

Gilbert, Jarrod. *Patched: The History of Gangs in New Zealand*. Auckland University Press, (2013).

Grimmer, Justin, and Brandon M. Stewart. 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.' *Political Analysis*, vol. 21, no. 3, (2013), pp. 267–297.,

Ivanusch, C. "Issue Competition in Parliamentary Speeches? A Computer-based Content Analysis of Legislative Debates in the Austrian Nationalrat" *Legislative Studies Quarterly*, Volume 49, Issue 1. <https://doi.org/10.1111/lsq.12421>, (2023)

New Zealand Parliament. *Hansard—Debate*. <https://www.parliament.nz/en/pb/hansard-debates/rhr/> Accessed 25 Mar. 2024.

NZ Legislation. "Gangs Legislation Amendment Bill". Government Bill <https://legislation.govt.nz/bill/government/2024/23-1> Accessed 25 Mar. 2024.

Radio New Zealand (RNZ) "Christopher Luxon wants gangs barred from promoting their lifestyle on social media" on *Radio New Zealand website* <https://www.rnz.co.nz/news/political/469138/christopher-luxon-wants-gangs-barred-from-promoting-their-lifestyle-on-social-media> 2022, Accessed 25 Mar. 2024.

Rychly, Pavel. A Lexicographer-Friendly Association Score. (2008). 2. vyd. Brno, RASLAN 2008. Brno: Masarykova Univerzita, p. 6-9. ISBN 978-80-210-4741-9.

Shaw, Richard. "Ch. 5 Government and the Westminster Tradition" *Government and Politics in Aotearoa and New Zealand*, edited by Janine Hayward, (2021), pp 278-323.

Williams et. al. "Corpus Building Project" University of Canterbury (2024).