

EXPLORING AUSTRALIA'S HANSARD (1901-2017)

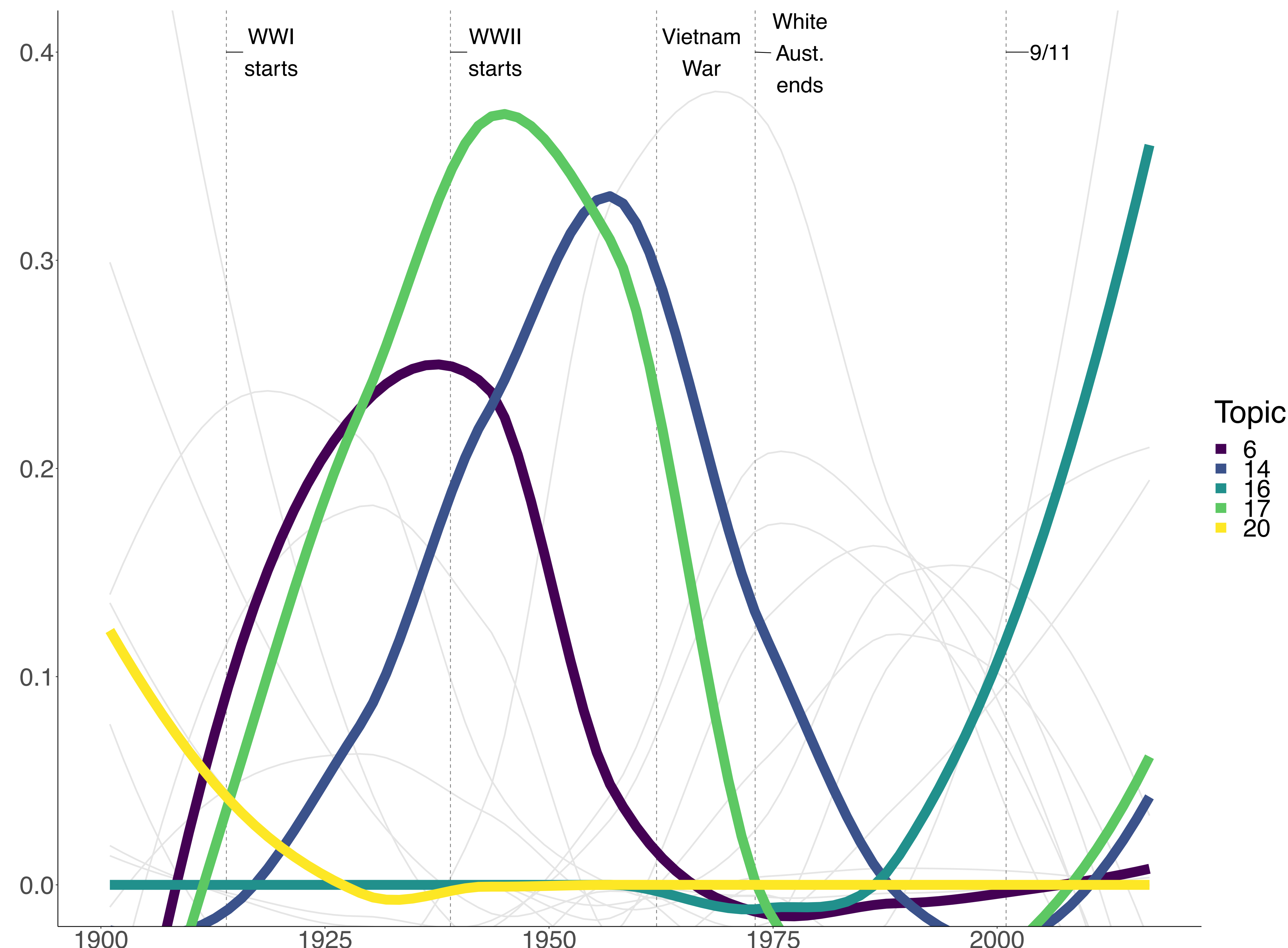
Rohan Alexander

Overview

I look at what was said in Australia's House of Representatives and Senate between 1901 and 2017. Using natural language processing I attribute centrality, topics and sentiment to each Hansard statement. I then associate these with various economic, political, and social factors. If a parliament in a modern liberal democracy is essentially just a 'talk shop' with little direct impact, then our thinking around this institution and its role may need to be more nuanced. Using text as an input to more usual analysis is becoming easier due to technology improvements and is an increasingly common aspect of many fields.

Data

The Australian Parliament House makes each day's Hansard available online as a PDF. I parsed these PDFs to convert them to text-based CSV files. The data are messy, but there are roughly 3 million statements made by 2,000 politicians over 15,000 house-days. Numbers, punctuation and common but less-informative words (such as 'Commonwealth') were removed and each word was stemmed, which allows, say, words and their plurals to be made indistinct. Topic modelling requires counts of each word by speaker and date. Event-driven changes and network effects can then be considered using standard modelling approaches.



House of Representatives.
Thursday, 9 May, 1901.

PROCLAMATION.
Honorable members assembled at half-past eleven a.m., in a chamber in the western annexe of the Exhibition-building, Melbourne, pursuant to the proclamation of His Excellency the Governor-General convening Parliament.

THE CLERK read the proclamation.

WRITING.
THE CLERK announced that he had received from the Secretary to the Prime Minister the following letter:—

Commonwealth of Australia,
Prime Minister,
Melbourne, 8th May, 1901.

Sir,
I have the honor, by direction of the Right Honorable the Prime Minister, to forward, herewith, the Writ of Election of 16 members to serve in the House of Representatives of the Commonwealth of Australia.

I have the honor to be,
Sir,
Your obedient servant,
ATLEE HUNT, Secretary.

The Clerk, House of Representatives, Melbourne.

The writs were introduced with ceremonies showing that the following gentlemen were duly elected for the States and districts set opposite their respective names, viz.—

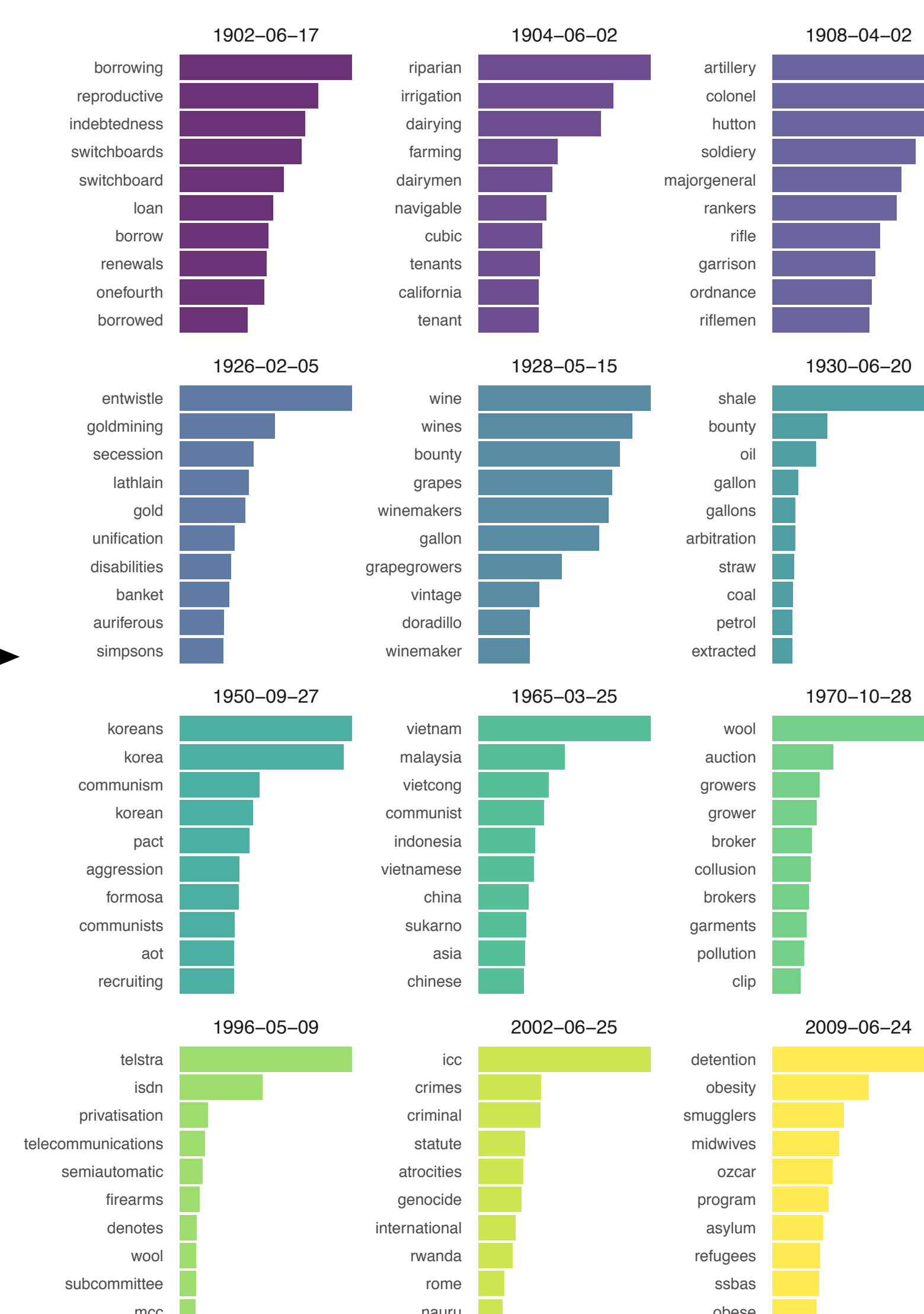
Frederick William Bamford, Esq., for the Electorate of Herbert, in the State of Queensland.

Right Hon. Edmund Barton, P.C., K.C., for the Electorate of Hunter, in the State of New South Wales.

Tigerton Leo Batchelor, Esq., for the State of South Australia.

Sir John Langdon Bonython, Kt., for the

Mr Briggs (Mayo) (10.07 Am)	I join with the member for Dobell, follow
Mr Melham	They have in the past; they will in relatio
Mr Briggs	They will in relation to decisions on rate:
Mr Melham (Banks) (10.14 Am)	I seek leave to speak without closing the
Mr Robb (Goldstein) (10.20 A...	I rise to speak on the Joint Standing Con
Mr Bradbury (Lindsay) (10.25...	I wish to add my congratulations to the e
The Deputy Speaker (Ms Ae ...	In my role as chair of committees I want
Mr Georganas (Hindmarsh) (...)	I seek leave to speak without closing the
Mr Georganas	Not many Australians are aware that Aus
Mr Irons (Swan) (10.56 Am)	As Deputy Chair of the House of Represe
Mr Hayes (Werriwa) (11.42 Am)	I seek leave to speak without closing the
Mr Hayes	I am pleased to table the Parliamentary J
The Deputy Speaker (Hon. Pe...	The honourable member should refer
Mr Hayes	Yes. I am glad I did not rush it and ment
Mr Wood (La Trobe) (11.49 A...	I thank the member for Werriwa, Chris H
The Deputy Speaker (Hon. Pe...	Before call the next item of business I
Mr Randall (Canning) (12.01 ...	I am pleased to rise in support of the Prc
The Deputy Speaker (Hon. Pe...	I thank the honourable member forences
Mr Oakeshott (Lyne) (12.40 P...	I rise today to appeal again to the Minist
Mr Fletcher (Bradfield) (12.5...	I rise to address the issue of value for m
Mr Neumann (Blair) (12.56 Pm)	I speak to congratulate the Lutherans of
Mr Combet	The Minister for Defence has provided th



Topic modelling

The latent Dirichlet allocation method (Blei, *et al.*, 2003) as implemented by the stm R package (Roberts, *et al.*, 2018) provides estimates of the topics of each document in a probabilistic sense. The key assumption behind LDA is that each document is generated by a speaker who decides the topics they would like to talk about in that document, and then chooses terms that are appropriate to those topics. A topic could be thought of as a collection of terms, and a document as a collection of topics. The collections are defined by probability distributions. The topics are not specified *ex ante*; they are an outcome of the method. Terms are not necessarily unique to a particular topic, and a document could be about more than one topic.

Next steps

Topics tend to change in response to events. It has been difficult to find factors led by Hansard, but more sophisticated models allowing correlation are promising.

Acknowledgements

Sincere thanks to my supervisor, John Tang, and to my panel: Zach Ward, Tim Hatton and Martine Mariotti. Thank you to Monica Alexander for helping in all aspects of this research.

✉ rohan.alexander@anu.edu.au
🐦 @RohanAlexander