

Words in the 45th Parliament: examining Senate debate through the frequency of keywords

Abstract

The Senate occupies an important role as a chamber of review; examining government legislation, debating public policy, and using the committee system to inquire into emerging or current issues. Debate, whether part of the legislative process or a part of the Senate's general business, is used to hold the executive accountable, to examine policy mechanisms and to raise the profile of issues. But are some issues spoken about in parliamentary debate more than others? Are some issues only debated infrequently? This paper examines debate in the Senate during the 45th Parliament using keyword searches to find the frequency with which particular words are used in the parliamentary sitting day. The paper analyses this information to show what issues are being debated more regularly, as well as what topics are not mentioned. Using this approach to analysing parliamentary debate raises questions about representation, particularly of minority issues. For instance, do Indigenous issues receive more mentions than the economy or jobs? Are environmental issues mentioned more frequently through the life of the parliament? Do some issues only receive a focus during a particular time, for example an awareness day or month? As well as answering these questions, this paper raises issues for future investigation. The paper's purpose is exploratory that is it aims to spark discussion around use of a different methodology for examining parliamentary debate.

Parameters of Senate debate and data mining

Debate, as Odgers' Australian Senate Practice notes, 'is one of the primary functions of the Senate; that of informing itself and the public by deliberation before decisions are made.'¹ The Senate's business is structured around making time for debate, whether in legislation or scrutiny of government policy.² Other parts of the Senate's daily business allow issues of public importance to be raised and debated, notices to be put to the Senate on public issues, and questions to be asked of the government.

Debate in the Senate chamber is also part of the Senate's function of effective scrutiny, described in Odgers' as 'to provide effective scrutiny of governments, and enable adequate expression of debate about policy and government programs'. In relation to the function of effective scrutiny, Odgers' notes that:

'[T]he significance of the Senate's role in these functions is that it is an elected and parliamentary forum. Other outlets for such debates in the community, for example, public conferences or print and electronic media, are not inherent institutions of democracy, though vital to it. As a parliamentary forum, moreover, the Senate is one place where a government can be, of right, questioned and obliged to answer. As such the Senate has been rightly seen as the safeguard of the Commonwealth.'³

The Standing Orders provide the parameters for business of the Senate to be conducted. Within the structure of the Standing Orders, the work of the Senate proceeds in a proportionate way, with time divided into the following categories: government business, business of the Senate, and general business. Changes to the order of business, the agenda of the Senate, must be made by the Senate itself. Within these categories are a number of procedurally defined areas in which debate can

¹ Odgers' Australian Senate Practice, 14th edition, Chapter 10.

² As set out in Standing Order 57 and related orders.

³ Odgers' Australian Senate Practice, 14th edition, Chapter 1.

occur. Most well-known is debate on legislation which encompasses the stages in the passage of a bill through the Senate from first reading through second reading speeches and committee of the whole consideration in detail to the third reading of the bill and its passage. Better known to the general public is Question Time during which questions without notice are asked of Government ministers by senators. The Standing Orders also set out other areas in which debate can occur including matters of public importance, senators' statements, taking note of answers provided in Question Time, motions for debate, and the daily adjournment debate. Within these areas, consideration of a wide variety of issues is possible.

Topics for debate arise through the government's program of legislation, progress of private senators' bills, or through debate in allocated times and forms; matters of public importance and urgency, question time, and motions. While the government generally has control over the legislative program, the other forms of debate are shaped by various influences. For example, matters of public importance are proposed by senators and selected by lot each morning.⁴ During question time, the content of the questions is decided by the parties themselves and generally each day has a thematic grouping of question topics.

The structured nature of debate in the Senate and the recording of all debates through Hansard transcripts allows for the use of data mining techniques to analyse Senate debate. There are a number of ways, using data mining techniques, in which topics debated in Parliament during sitting can be identified. For example, tracking daily legislative programs gives a broad overview of what has come before a chamber. The Journals of the Senate record issues that are raised through motions and the topics of matters of public importance can be tracked day to day in this way too. Senate Statsnet tracks questions without notice asked each sitting day. This provides a break down by party, senator and title of the question through which it is possible to look at the topics being raised during Question Time.

The data mining approach used in this paper is word count frequency. This technique uses word searches based on keywords and identifies patterns using the frequency of keywords and the areas of debate in which clusters of words appear. Using this technique to track topics debated in a particular chamber offers a different way to examine parliamentary debate. Focusing on the number of times a keyword appears in a day's transcript can reveal the frequency with which a topic is raised, unconnected to whether the debate was legislative or during another part of Senate business. Comparisons are possible between topics in relation to the frequency they are discussed and whether that frequency changes with any external event, for example media focus or significant public debate.

Methodology of word frequency searches for Hansard transcripts from the 45th Parliament

While the database of ParInfo is available for searching keywords and phrases, the methodology used for this paper relied on examining individual Hansard transcripts for each Senate sitting day in the 45th Parliament. To conduct the word searches, each Hansard PDF file was converted to Microsoft Word and then the word search function was used to identify the number of times a keyword appeared in the transcript. Word count totals were compiled in an Excel spreadsheet allowing patterns to be identified across the months and years of the 45th Parliament.

Microsoft Word format was chosen because it allows flexibility in data gathering. In Microsoft Word format the overall number of appearances of the keyword can be found through a word search. Clusters where the keyword appears in the text can be checked using the same function. This latter

⁴ Standing Order 75.

was assessed by reading through the Hansard to see clusters of a keyword and the section in which they appeared. The ability to pinpoint word frequency in a single day of debate was useful in identifying patterns based on which type of debate issues appeared in most frequently.

To create data for comparison, eight keywords were searched for in all Hansard transcripts of the 45th Parliament: *LGBT*, *environment*, *climate*, *Indigenous*, *health*, *education*, *economy*, and *job*. These words were chosen either to represent minority groups (*LGBT*, *Indigenous*), policy areas (*environment*, *health*, *education*, *climate*), or words commonly associated with politics (*economy*, *job*). The keywords were also chosen because their usage is more likely to be directly related to their topic. For example, *health* as used in parliamentary debate is more likely to refer health policy. Similarly *education* is likely to refer to education policy. By looking through the individual transcripts in which *climate* appears, it was found that majority of the uses of *climate* refer to climate change. However, given the use of Microsoft Word documents for word searches, it was not possible to search composite terms like "climate change" or "same sex marriage", hence the use of the acronym *LGBT* to track debate on same-sex marriage and related issues. It could be argued that searching "marriage" would be a way in which to track the same-sex marriage debate, but because the word can occur commonly in general usage as well as specifically in relation to same-sex marriage, searching this way would require additional keyword searches to provide context around usages of "marriage".

Identifying word frequency patterns in debate

Average word counts for keywords in the 45th Parliament show interesting patterns, as in Table 1. Within this group of keywords *health* was the most talked about topic of the 45th Parliament in the Senate, followed by *job*. Despite same-sex marriage being a major issue in 2016 and 2017, with same-sex marriage legalisation being before the Senate, *LGBT* is the least mentioned issue on a keyword search. Interestingly, the rates of mention of *environment*, *climate* and *Indigenous* are all low in comparison to *health*, *education* and *job*. At first glance, average keyword counts give a picture of the frequency with which topics were debated in the Senate during the 45th Parliament; this is the first stage of analysis of the data collected. To understand more about which areas of Senate debate keywords and their associated issues were being used in, it is necessary to further analyse the individual day word searches. The following discussion will centre on the keywords *health*, *climate* and *Indigenous* to examine how patterns of word frequency can be found by analysing Hansard transcripts for one Parliamentary year – 2018.

Table 1 – average word mentions in the 45th Parliament

Year	LGBT	Environment	Climate	Indigenous	Health	Education	Economy	Job
2016	9.43	37	22.39	31.60	117.47	56.73	36.17	87.34
2017	7.43	46	23.30	33.18	82.38	79.45	30.32	99.4
2018	6.79	48.34	26.62	39.22	112.81	65.48	46.46	124.37
2019	4.2	73.8	68.4	56.8	92	65.8	35.8	81

Health, *climate* and *Indigenous* have been chosen because each reflects a key policy area. They also provide a variety of examples of the ways in which topics can be tracked through different areas of debate. Further, in the case of *Indigenous*, the patterns observed can be linked to events external to the Parliament.

To look at patterns for these keywords, instances in which the word frequency occurs well above average have been chosen as a focus. Tables 2 to 4 list these examples for the keywords, noting the areas of debate in which the keyword frequency was highest on specific days. This data is gathered

by searching the specific day's transcript to see where clusters of keywords are most frequent and matching this to the part of the parliamentary debate where the clusters appear.

What the examples show is that on these days' debate on these topics increased markedly whether because of current legislation or because of deliberate inclusion of the topics in time allocated for debate. The type of debate is of interest as topics included in each type of debate are determined differently. For example, questions without notice are decided by the parties of those senators asking the questions, whereas issues for debate in a matter of public importance are submitted for a daily ballot. Examining the forms of debate in which an issue is discussed, as well as the frequency of the issue, may provide an indication of the importance of the issue and its priority among senators.

Legislative debate centres on the piece of legislation under consideration. Legislative debate (initiated by a minister) is government business and is given precedence over general business (initiated by other senators). The time set aside for debate on government business in the senate (primarily government legislation) helps to explain why pieces of legislation cause a focus on a debate topic for extended periods of time, in comparison with the shorter amounts of time allocated to other forms of debate. In legislative debate, it is generally the second reading stage which is the most significant and lengthiest. It is this part of consideration of a bill which is likely to contain most debate about the wider policy related to the bill; with committee of the whole consideration of the bill examining the detail of the bill itself.

Word frequency in legislative debate

In regards to the keyword *health*, in 2018 the lowest number of mentions of *health* is 28 on 13 September. The highest is 519 on 10 September and the average is 112.8. In 2018 *health* was most likely to occur in a legislative context, with six pieces of legislation providing a large amount of debate time devoted to health policy. A review of the Hansard for 10 September shows that on this day *health* was mentioned four times more than the average for 2018, was due to three health related pieces of legislation being debated that day.

Aside from legislative debate, *health* also appeared in questions without notice and motions to take note of answers, but it was less frequent in these areas of debate than in legislative debate. Interestingly, out of the 58 Senate sitting days in 2018 (excluding Senate Estimates), questions related to health were asked rarely. *Health* was more likely to be debated during a matter of public importance than raised in Question Time: three matter of public importance debates dealing with health occurred in 2018. Table 2 shows the division of debate for days when *health* had a frequency above average.

Table 2 – occurrences of *health* in 2018 and forms of debate

2018	Items at which <i>health</i> was mentioned most frequently
5 February – 171 mentions	Question without notice and taking note of answers
12 February – 164 mentions	Debate on the National Health Amendment (Pharmaceutical Benefits—Budget and Other Measures) Bill 2017
19 March – 152 mentions	Debate on the National Health Amendment (Pharmaceutical Benefits—Budget and Other Measures) Bill 2017
8 May – 154 mentions	Matter of public importance debate and adjournment debate
21 June – 168 mentions	Debate on the Health Legislation Amendment (Improved Medicare Compliance and Other Measures) Bill 2018
25 June – 230 mentions	Matter of public importance debate was on health care
26 June – 168 mentions	Questions without notice
10 Sept – 519 mentions	Debate on three pieces of legislation – Private Health Insurance Legislation Amendment Bill 2018, A New Tax System (Medicare Levy Surcharge – Fringe Benefits) Amendment (Excess Levels for Private Health Instance Policies) Bill 2018, and Medicare Levy Amendment (Excess Levels for Private Health Insurance Policies) Bill 2018
12 Nov – 194 mentions	Matter of public importance debate
14 Nov – 357 mentions	Debate on the My Health Records Amendment (Strengthening Privacy) Bill 2018
15 Nov – 250 mentions	Debate on the My Health Records Amendment (Strengthening Privacy) Bill 2018

In comparison with other keywords, *health* appears much more frequently in the 45th Parliament. In 2018 it is second only to *job* (at an average of 124.37 mentions) and the next most frequent keyword is *education* at an average of 65.48 mentions. Examining the Hansard transcripts for days when mentions of *health* are above average, it is clear that the high word frequency is due to health-related legislation being debated. This demonstrates that on an average day in 2018, *health* was less likely to arise in regards to a question without notice or another form of debate, than it was to be discussed by the Senate in the context of specific legislation. Such an observation is interesting because it shows that *health* as a policy issue in 2018 was more likely to be linked with specific government policy than it was to be debated more broadly.

Appearance of a keyword and issue in legislative debate and rarely in other forms of debate may account for the low frequency of *LGBT* during the 45th Parliament. Even though there was a large amount of public debate about same-sex marriage, the amount of time the Senate spent on debate about same-sex marriage was relatively low as it was restricted to consideration of legislation relating to the same-sex marriage plebiscite, postal survey and legalisation.

Comparing word frequency across forms of debate

The keyword *Indigenous* has an average of 39.2 occurrences in 2018, with a peak of 244 on 12 February 2018 and lowest count of 11 (on two days, 20 June and 17 October). Legislative debate accounted for a large number of occurrences of *Indigenous*, although the frequency was higher in other forms of debate. In comparison with *health* and *climate*, *Indigenous* appeared in no matter of public importance or urgency debates in the dates reviewed. Appearance in other forms of debate, for example adjournment debate, was also rare for *Indigenous* as a term. Similar to *health* however, legislative debate could cause a significant increase in the number of mentions for *Indigenous*. Unlike the keyword *health* and similarly to the keyword *climate*, *Indigenous* also appeared in debate

about legislation that did not have the word in its title. For example there were mentions of *Indigenous* during debate on Treasury Laws Amendment (2017 Measures No. 5) Bill 2017 and the ASIC Supervisory Cost Recovery Levy Amendment Bill 2017.

Interestingly, usage of the keyword *Indigenous* occurred more frequently in Question Time than *health* or *climate*, with most days of above average frequency involving questions without notice directly related to Indigenous affairs. This pattern is of interest because the purpose of Question Time debate is to allow the Government to be questioned by senators about its policies. It is the most often televised part of the Senate's proceedings and it is the part of the parliamentary day which is most likely to be familiar to members of the general public. Another feature of Question Time is that political parties, most notably the opposition and the Government, are in control of the topics of questions. The decision to raise a topic in the form of a question to Government is deliberate and makes the issue a focus at a key point in the parliamentary sitting day.

Table 3 – occurrences of *Indigenous* in 2018 and forms of debate

2018	Items at which <i>Indigenous</i> was mentioned most frequently
12 Feb – 244 mentions	Questions without notice; motions to take note of answers; motions for debate; debate on the Social Services Legislation Amendment (Cashless Debit Card) Bill 2017
13 Feb – 118 mentions	Questions without notice; motions for debate; adjournment debate; debate on the Prime Minister and Cabinet Legislation Amendment (2017 Measures No. 1) Bill
20 March – 135 mentions	Questions without notice; order for the production of documents relating to Indigenous housing
28 March – 87 mentions	Questions without notice; some mentions during debate on Treasury Laws Amendment (2017 Measures No. 5) Bill 2017 and the ASIC Supervisory Cost Recovery Levy Amendment Bill 2017
14 August – 56 mentions	Questions without notice; some mentions during debate on Restoring Territory Rights (Assisted Suicide Legislation) Bill 2015
16 October – 53 mentions	Questions without notice; motions to take note of answers
18 October – 57 mentions	Debate on the Social Security Legislation Amendment (Community Development Program) Bill 2018
12 November – 76 mentions	Debate on three pieces of legislation – Aboriginal and Torres Strait Islander Amendment (Indigenous Land Corporation) Bill 2018, Aboriginal and Torres Strait Islander Land and Sea Future Fund Bill 2018, and Aboriginal and Torres Strait Islander Land and Sea Future Fund (Consequential Amendments) Bill 2018; mentions during documents and motions for debate
15 November – 73 mentions	Questions without notice; additional answers; order for the production of documents relating to the Indigenous Advancement Strategy
28 Nov – 122 mentions	Debate on three pieces of legislation – Aboriginal and Torres Strait Islander Amendment (Indigenous Land Corporation) Bill 2018, Aboriginal and Torres Strait Islander Land and Sea Future Fund Bill 2018, and Aboriginal and Torres Strait Islander Land and Sea Future Fund (Consequential Amendments) Bill 2018; statements by senators; and related topics in questions without notice

Influence of significant public events

As Table 3 shows, the major increase in word frequency for *Indigenous* was on 12 and 13 February 2018. While some of the frequency on 12 February can be attributed to debate of the Social Services Legislation Amendment (Cashless Debit Card) Bill 2017, the frequency on 12 and 13 February is six and three times, respectively, the average of the word frequency in 2018. One factor may account for this significant increase: 13 February was the ten year anniversary of the National Apology to the Stolen Generations and Indigenous people of Australia. It was also around this date that the annual Closing the Gap report was presented. As a result of this significant event, there was more media and parliamentary focus on Indigenous affairs.

Word frequency can be used to demonstrate the effect that a significant day like 13 February has on debate in the Senate. In comparison to other days of above average frequency for *Indigenous*, only one other day stands out – 20 March on which there were three times the average mentions of *Indigenous*. The major increase in frequency on this day can be attributed to questions without notice and an order for production of documents relating to Indigenous housing. The order for production of documents related to correspondence between the then Minister, Senator the Hon Nigel Scullion, and his state and territory counterparts on the National Partnership Agreement on Remote Housing.⁵ There was significant debate around the order, as there had been media interest in the funding for remote housing. The continuation of media concentration on the issue⁶ most likely contributed to increased frequency of *Indigenous* on other subsequent sitting days.

Comparisons with keyword searches in Parlinfo

For *climate*, three mentions is the lowest point (12 February, 26 March, and 17 September) while the peak frequency is 111 on 4 December. The average mentions of *climate* is 26.6 for 2018. In comparison to *health*, *climate*, which can be used to track debate on climate change, appears much less. Mentions of *climate* only passed 100 on two occasions in 2018, in comparison with *health* having an average of 112.8 mentions in 2018. This may indicate the contrasting priority that the two issues were given in Senate debate, but equally it could indicate that *health* had more legislative debate and therefore greater debate time overall.

Climate was mentioned in conjunction with less legislative debate. However *climate* appears in legislative debate across a variety of legislative areas from legislation relating to the Great Barrier Reef to the Pacific Agreement on Closer Economic Relations Plus in international relations. *Climate* appeared in more questions on notice and arose frequently in the period of motions to take note of answers. During this period of debate, held after Question Time, senators are able to speak at length regarding the answers provided by Ministers in Question Time. *Climate* also arose during a matter of public importance and two matters of urgency during 2018. Table 4 shows the division of debate for transcripts when *climate* had an above average number of occurrences.

⁵ 'Future of remote Indigenous housing strategy uncertain', Nakari Thorpe, NITV News, 21 March 2018, <https://www.sbs.com.au/nitv/nitv-news/article/2018/03/21/future-remote-indigenous-housing-strategy-uncertain>

⁶ An example of continued coverage about the National Partnership Agreement on Remote Housing is 'Scullion hits back over remote housing funding dispute', Tom Zaunmayr and Peter De Kruijff, Pilbara News, 12 July 2018, <https://thewest.com.au/news/pilbara-news/scullion-hits-back-over-remote-housing-funding-dispute-ng-b88893805z>

Table 4 – occurrences of *climate* in 2018 and forms of debate

2018	Items at which <i>climate</i> was mentioned most frequently
8 Feb – 61 mentions	Debate on the Great Barrier Reef Marine Park Amendment (Authority Governance and Other Matters) Bill 2017
19 March – 53 mentions	Debate on the Coal-Fired Power Funding Prohibition Bill 2017; matter of public importance debate
13 August – 90 mentions	Questions without notice; motions take note of answers; matter of public importance debate on drought
23 August - 69 mentions	Questions without notice; motions for debate
10 September – 62 mentions	Debate on the Coal-Fired Power Funding Prohibition Bill 2017; questions without notice
13 September – 50 mentions	Debate on two bills – Customs Amendment (Pacific Agreement on Closer Economic Relations Plus Implementation) Bill 2018, Customs Tariff Amendment (Pacific Agreement on Closer Economic Relations Plus Implementation) Bill 2018; motions for debate
28 November –66 mentions	Motions to take note of answers; motions for debate
4 December – 111 mentions	Questions without notice; motions to take note of answers; order for production of documents; matter of urgency debate regarding climate change
5 December – 104 mentions	Statements by senators; questions without notice; motions to take note of answers; debate on the Galilee Basin (Coal Prohibition) Bill 2018; motions for debate; matter of urgency debate regarding the mining industry

In comparison to *health* and *Indigenous*, the frequency of *climate* spans different areas of debate and does not necessarily arise in one particular area of Senate debate more than others. This may indicate that debate on climate change arises less by design, that is, political parties deciding to focus questions without notice or debate on the issue, and more by association with other topics.

Comparison of keyword searches in Hansard with the same keyword search in Parlinfo media for 2018 gives an indication of how Senate debate on the topic of climate change progressed relative to the public debate. Some context considerations are necessary in this comparison. For example in October 2018 there were four sitting days in the month and in May 2018 there were three, representing less time for debate. Such a consideration may account for the low frequency of *climate* in these months in comparison to the increasing frequency in Parlinfo data for these months. With these considerations in mind, the comparative data in Table 5 is interesting in that it shows an increase in the frequency of *climate* in media alongside an increase in frequency in Senate debate. The question could be asked as to whether parliamentary debate is being influenced by media or whether media reporting is increasing because of the prominence of the issue in parliamentary debate.

Table 5 – comparison of *climate* word frequency in Hansard with word frequency in Parlinfo media database

	<i>Climate</i> mentions in Senate Hansard, 2018	<i>Climate</i> search results for media through Parlinfo search
February	137	516
March	138	630
May	46	619
June	91	674
August	356	1717
September	211	966
October	83	1723
November	235	1286
December	237	1172

Word frequency analysis – the potential for future research

Using keyword searches to analyse Senate Hansard transcripts produces a different way of viewing Senate debate. This method allows issues to be tracked through Senate debate; and to identify what might cause the Senate to focus on a particular issue. The findings are interesting. For example, of the keywords used for this paper *health* was used more frequently in the 45th Parliament than other words. Frequency of *LGBT* was low in comparison to other words even with increased public debate on same-sex marriage, which can be attributed to the Senate debating the issue only when relevant legislation was scheduled for debate. Majority of instances of *Indigenous* were during questions without notice or motions to take note of answers. In comparison, *health* had less frequency in Question Time and more usage during legislative debate. The patterns shown in the keyword searches give a picture of how Senate debate proceeds, what issues occur more frequently, and in what forms of debate.

Eight keywords have been used for the purposes of this paper, but the method is applicable to any number of other keywords. With different methods of searching, that is, other than using individual Microsoft Word documents, more variety in keywords and phrases may be possible. Using Microsoft Word documents and conducting individual keyword searches was useful for this paper because it allowed search results for a specific day to be interrogated through detailed examination of the transcript. Other Senate transcripts could also be searched in this way, for example committee hearing transcripts or Senate estimates proceedings. The same method could also be applied to Hansard transcripts for the House of Representatives, allowing a comparison between the two chambers for particular periods of time.

Expanding the keyword search could also be useful in determining other words used in debate. For example, while *LGBT* was used to track debate on same-sex marriage, other words could be used to examine the tone of debate. This would reveal the language used around a topic as well as the frequency with which the topic is discussed. The same approach could be used to further examine data gathered by keyword searching *health*. For example, do these debates focus on hospitals or on primary care? Focusing on surrounding words could also be useful in examining particular forms of debate as language usage may vary when topics are discussed in legislative debate in comparison to when they are debated in a matter of public importance.

A wide variety of statistics are available about the work of the Senate, from detailed lists of questions without notice topics to the length of legislative debates. Keyword searches of Hansard

transcripts add another form of analysis of Senate debate, one which can be used to identify patterns of topics as well as examination of the words used around a particular topic. The use of this method of analysis for Senate Hansard transcripts has a wide range of possibilities for future research.